GRAPHISTRY

Scaling investigations with massive visual graph analytics

NVIDIA GTC 2019

info@graphistry.com

GRAPHISTRY

Team • Berkeley, Netflix, Google, Fortinet, …

Users • Security, anti-fraud, counter-trafficking, NOC, trouble shooting, …

OSS • goai, Apache Arrow, RAPIDS, RxJS, FALCOR

NVIDIA, IQT IN-Q-TEL, Bloomberg BETA

DEPARTMENT OF DEFENSE UNITED STATES OF AMERICA, FORTUNE 500

COORDINATED ATTACK

A French security expert reveals how trolls backing the BJP and Congress fuelled #BoycottSurfExcel

Republican Senator Conversations on Twitter: Bot or Not?

Journey to 100X'ing Investigations

- Why big graphs
- "No code" for ML & automation
- GPUs: GOAI in Graphistry Engine 2.0

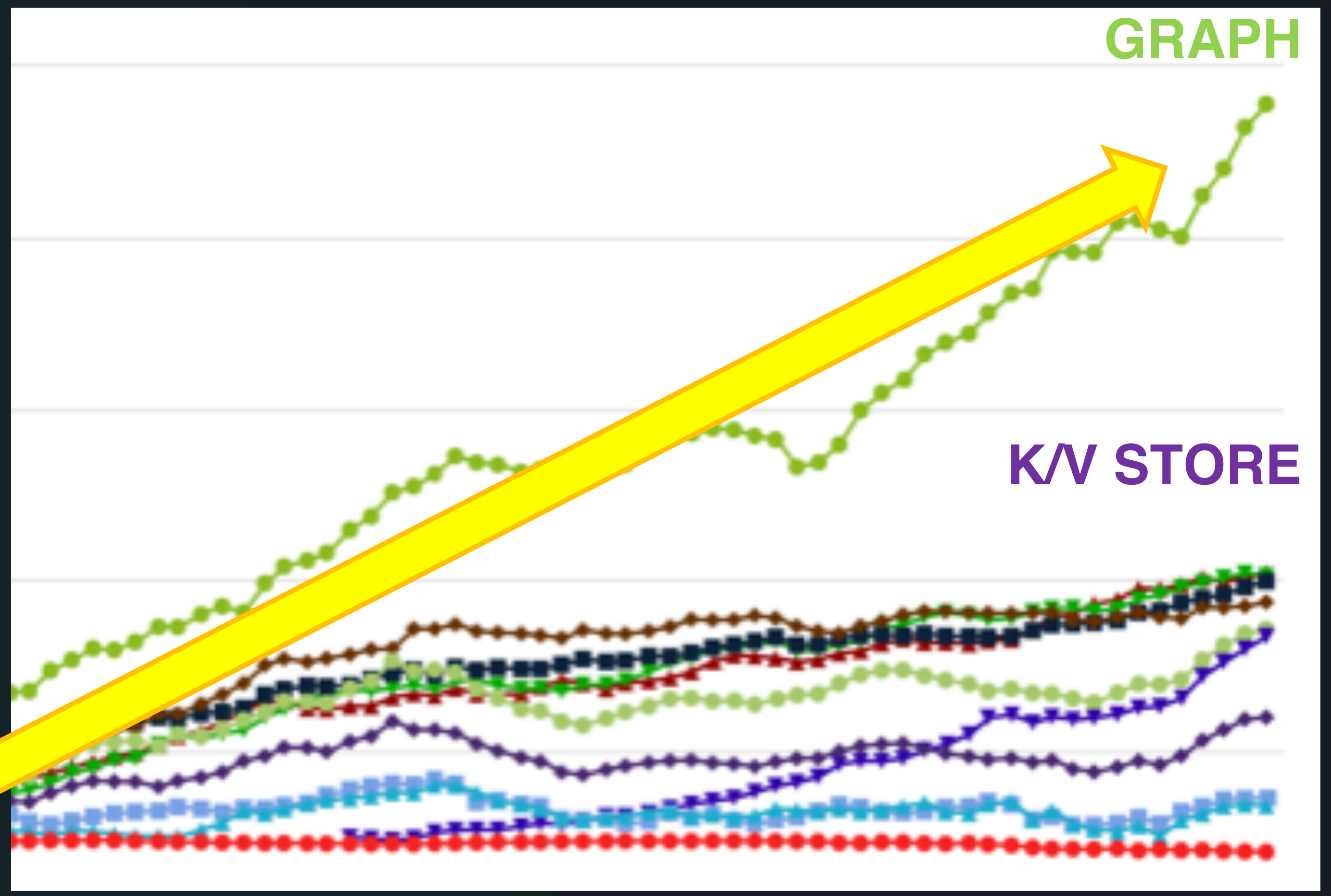
The next 5 years will be filled with graph analytics

Gartner

Top 5 CAGR in Data Tech

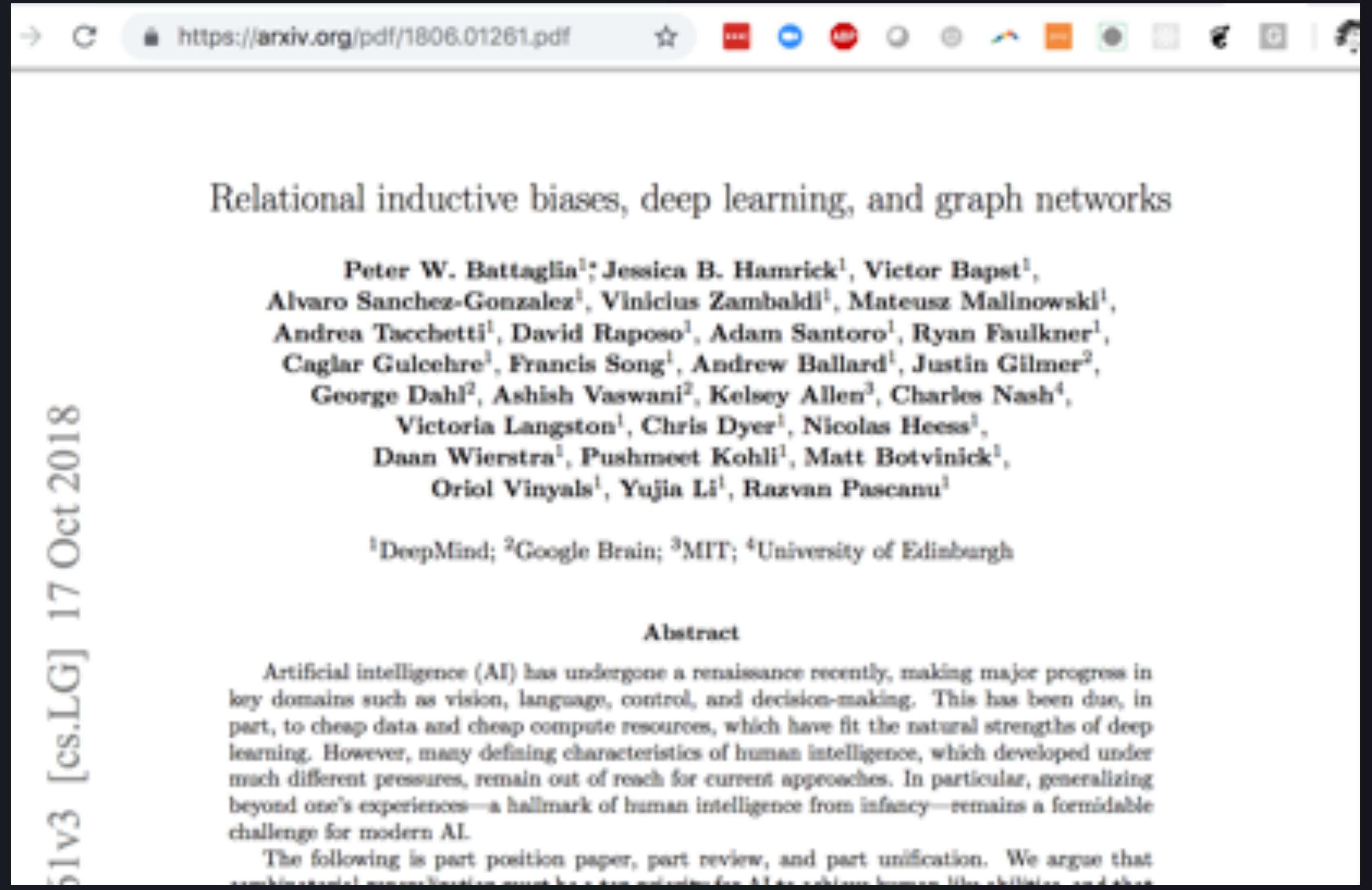
DB-ENGINES

Fastest Growing DB Category



Google DeepMind

AI Manifesto preprint: Oct 2018


https://arxiv.org/pdf/1806.01261.pdf

Relational inductive biases, deep learning, and graph networks

Peter W. Battaglia[1], Jessica B. Hamrick[1], Victor Bapst[1], Alvaro Sanchez-Gonzalez[1], Vinicius Zambaldi[1], Mateusz Malinowski[1], Andrea Tacchetti[1], David Raposo[1], Adam Santoro[1], Ryan Faulkner[1], Caglar Gulcehre[1], Francis Song[1], Andrew Ballard[1], Justin Gilmer[2], George Dahl[2], Ashish Vaswani[2], Kelsey Allen[3], Charles Nash[4], Victoria Langston[1], Chris Dyer[1], Nicolas Heess[1], Daan Wierstra[1], Pushmeet Kohli[1], Matt Botvinick[1], Oriol Vinyals[1], Yujia Li[1], Razvan Pascanu[1]

[1]DeepMind; [2]Google Brain; [3]MIT; [4]University of Edinburgh

Abstract

Artificial intelligence (AI) has undergone a renaissance recently, making major progress in key domains such as vision, language, control, and decision-making. This has been due, in part, to cheap data and cheap compute resources, which have fit the natural strengths of deep learning. However, many defining characteristics of human intelligence, which developed under much different pressures, remain out of reach for current approaches. In particular, generalizing beyond one's experiences—a hallmark of human intelligence from infancy—remains a formidable challenge for modern AI.

The following is part position paper, part review, and part unification. We argue that

[cs.LG] 17 Oct 2018

Early Winners Need Disrupting


Search

Bloomberg

Technology

Palantir Weighs IPO Valued at $41 Billion, WSJ Says

By Molly Schuetz

October 18, 2018, 6:15 AM PDT

Virtual Graph is Emerging

GraphQL

dremio

IFTTT

zapier

Microsoft Graph

Graphistry turns enterprise data into graph apps & insights

Everything important is now mirrored in logs

SECURITY, FRAUD, OPERATIONS, ...

- 1K – 1M devices
- 1K – 1B users
- Digital activities: Payments, logins, clicks, …

Analysts love graphs

- See alerts in context, not isolation
- Prioritize according to risk
- Leverage data without overwhelming the analyst

MATT SWANN

Microsoft

GRAPHISTRY

Warmup: The first Blockchain APT (Advanced Persistent Threat)

- Ethereum theft: $1M / week

- Automatically map out 20K+ flagged transactions: Entities, behavior, timeline

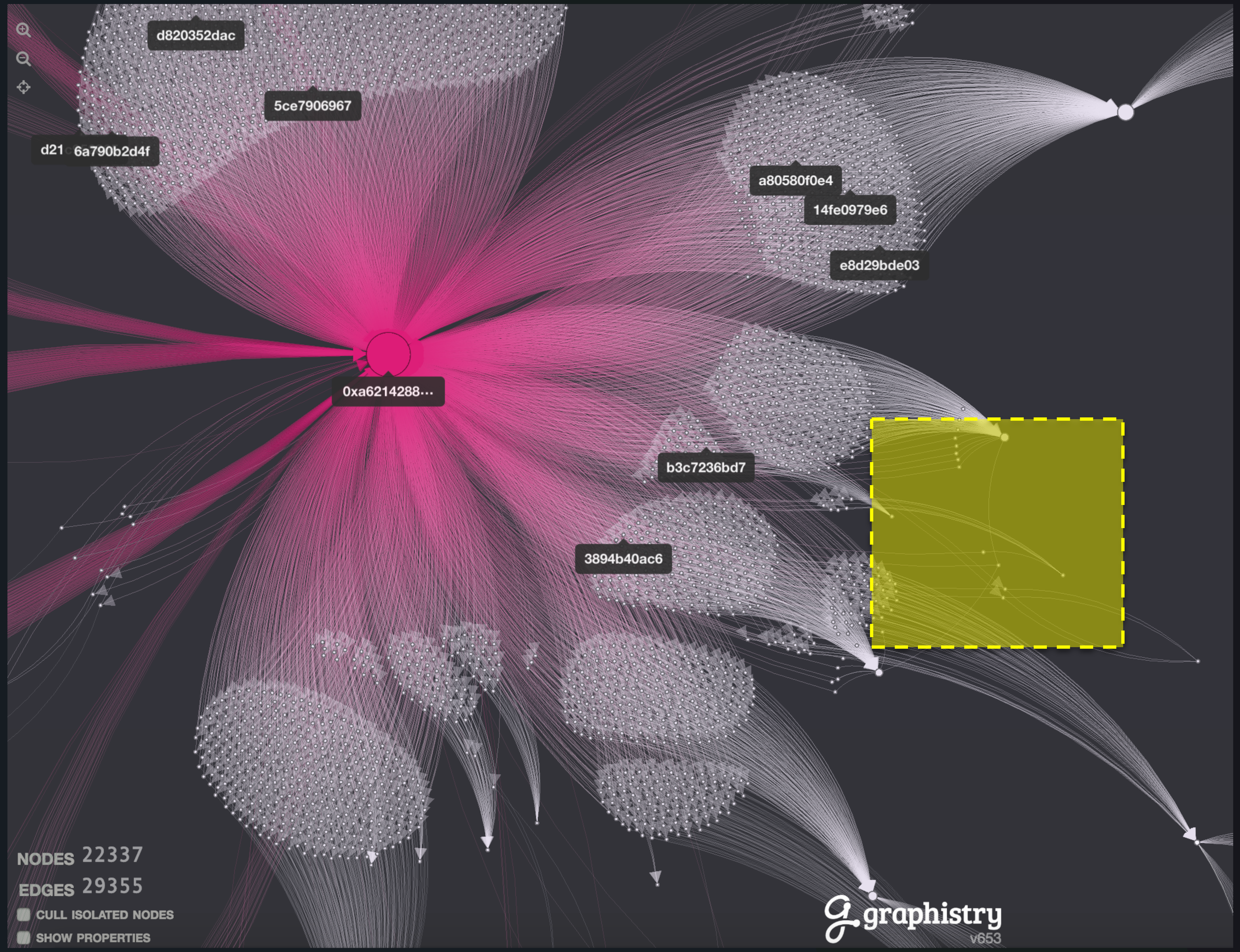
区块链技术与应用
区块链产业服务平台

头条 快讯 链闻

硅谷AI区块链安全公司AnChain.ai与数据可视化公司Graphistry达成合作

2018-08-29

近日，硅谷领先的区块链安全服务公司AnChain.ai与数据可视化公司Graphistry达成合作。 早在本月初，AnChain.ai就通过态势感知发现了隐藏在以太坊游戏FOMO3D和Last Winner背后的黑客团伙 BAPT-F3D和BAPT-LW20，并通过与Graphistry的合作，视觉化还原了黑客攻击的全过程。该黑客团伙通过智能合约漏洞，发动攻击，仅仅4天就获利5194ETH，约合1200W人民币。 AnChain.ai与Graphistry的此次合作调查证实了区块链领域高级持续性攻击（即BAPT，由AnChain.ai发现并命名）的存在，双方将发挥各自在人工智能、区块链安全、数据挖掘、数据可视化等领域的专长，最大限度的保障 Dapp 开发者、交易所及区块链生态的安全与透明。



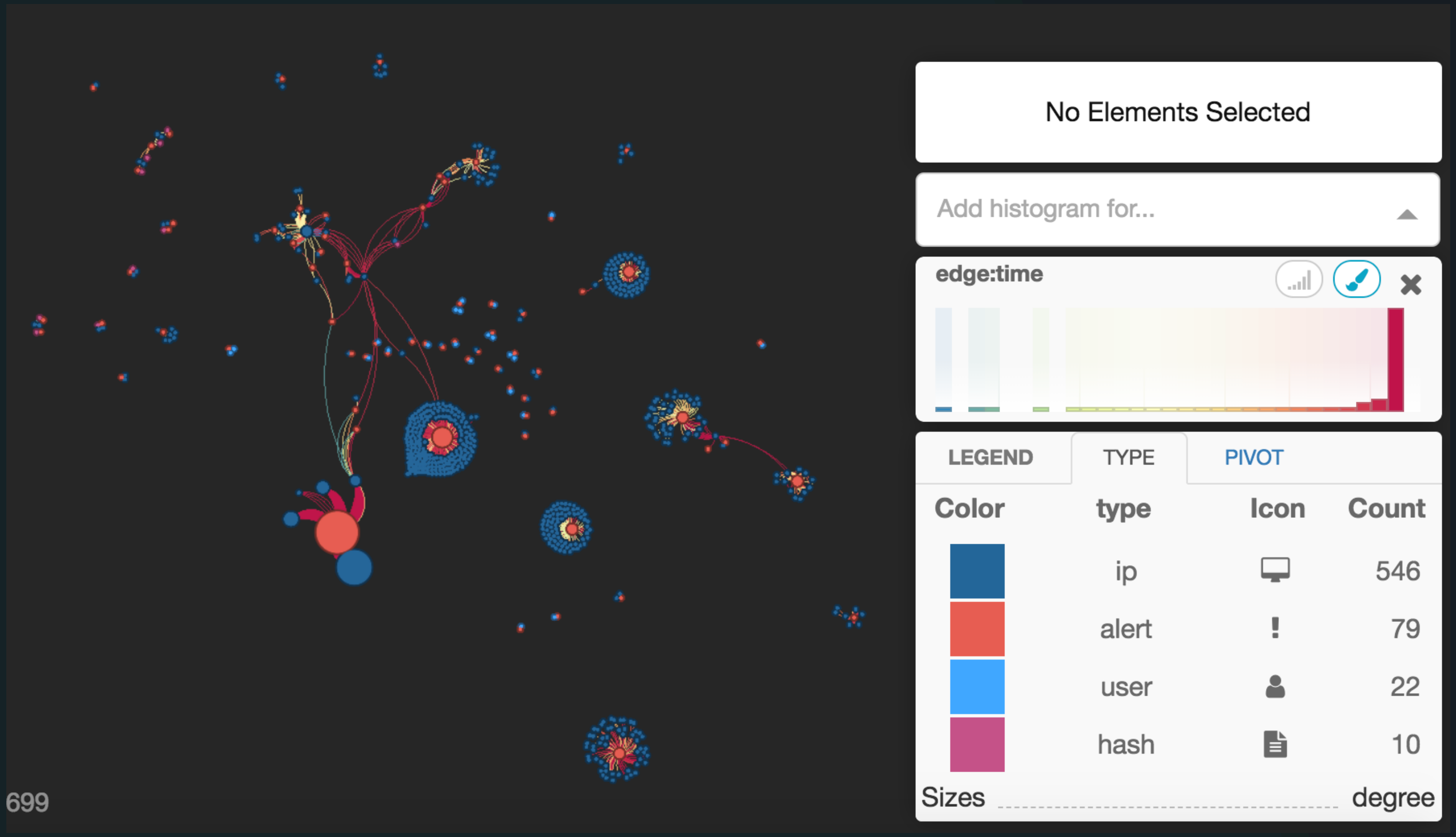
Rich logs have the answers. Expose them with the hypergraph trick!

Security (CIM, CEF, ..): 400+ columns over rich JSON values

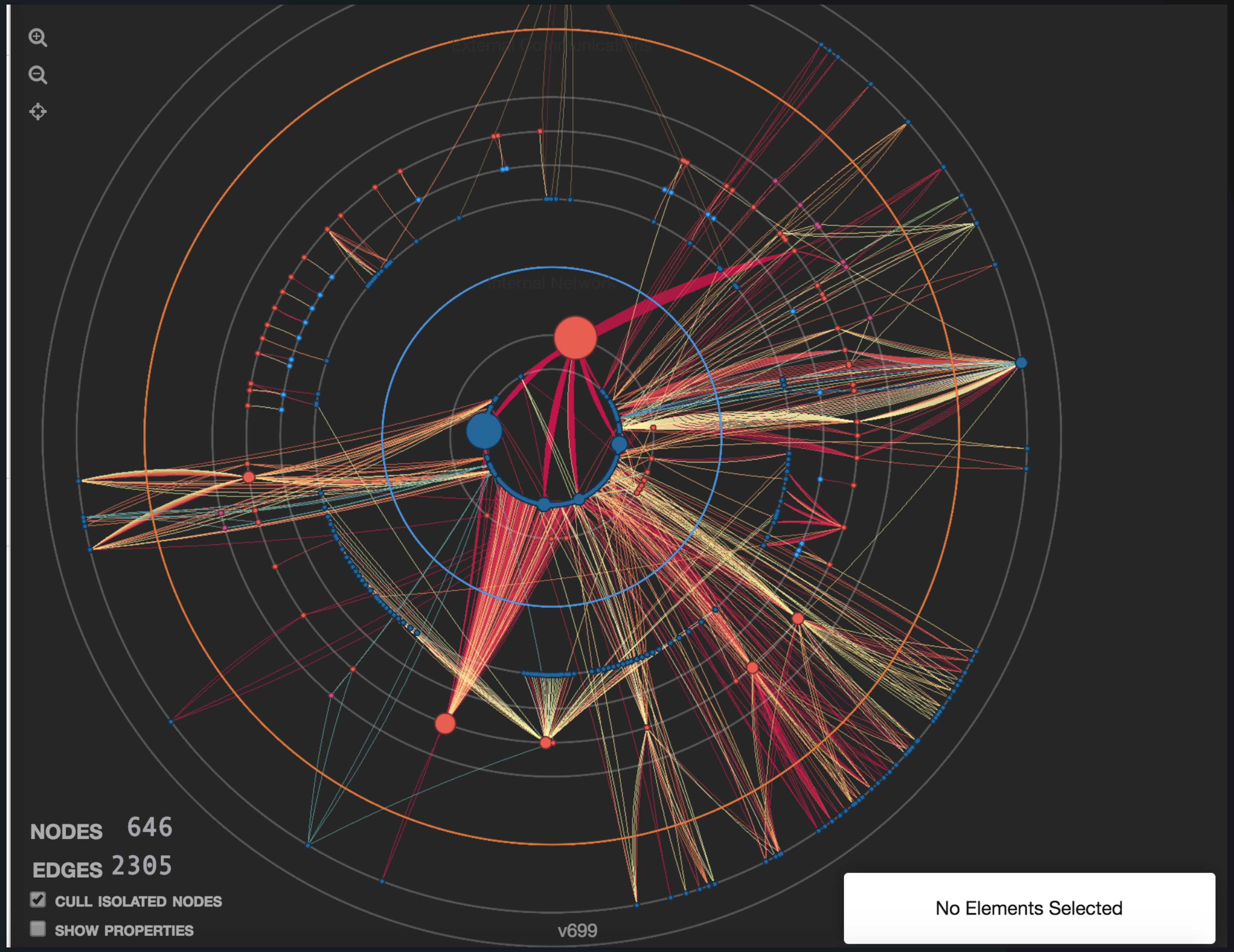
Hypergraphs: 1) Extract unique entities 2) Link if ever co-occurred

Demo: Threat Intelligence for Malware Campaigns

$B Slide: Enterprise event analytics



events grouped into incidents



security perimeter crossings

events → grouped incidents → contextualized alerts → actions

No code

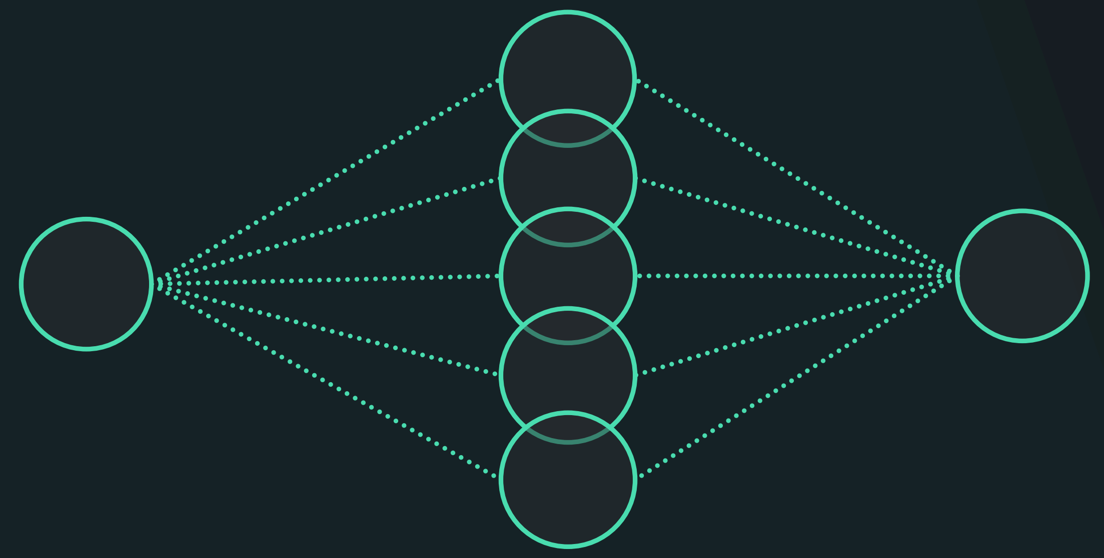
INVESTIGATIONS
Search
OODA Loop
Runbook
Data Wrangling
...

- Query
- Correlate
- Decide
- ... repeat

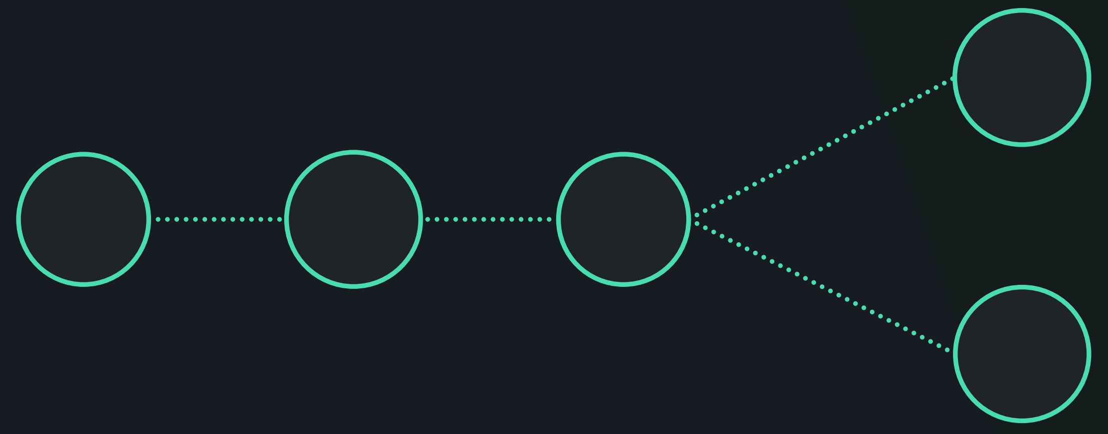
➔ Appropriate actions

No code -- Graph makes AI explainable

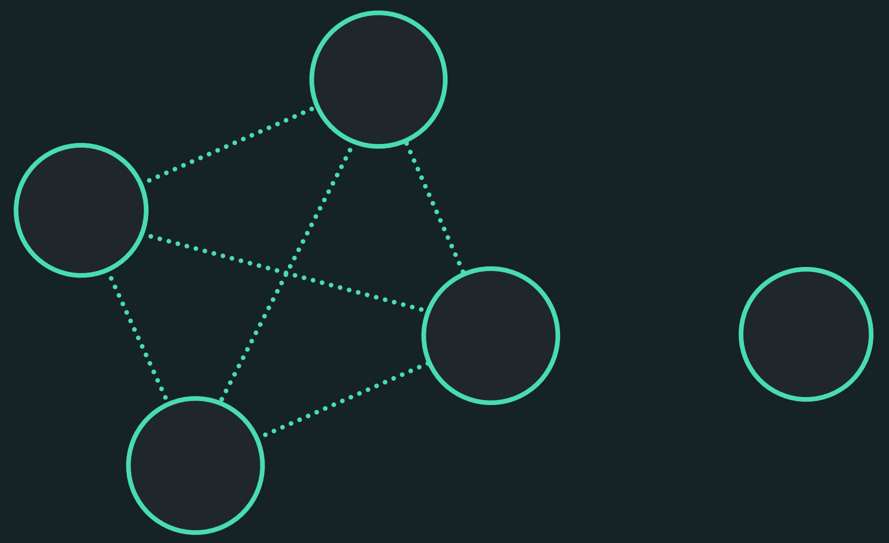
Scoping



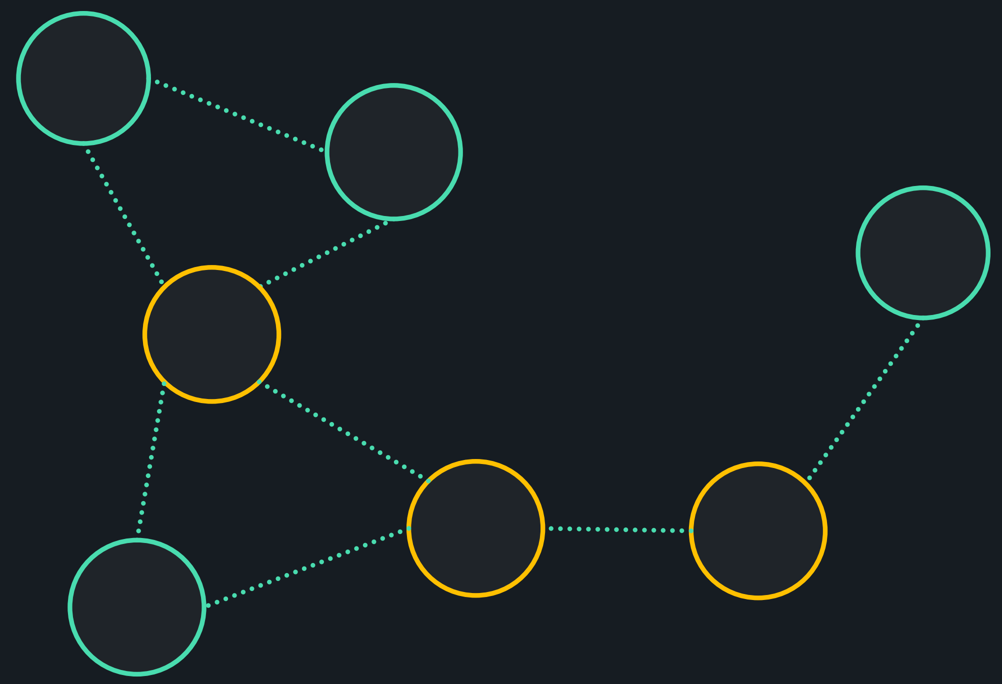
Progression & Behavior



Patterns & Outliers



Influence & Critical Players

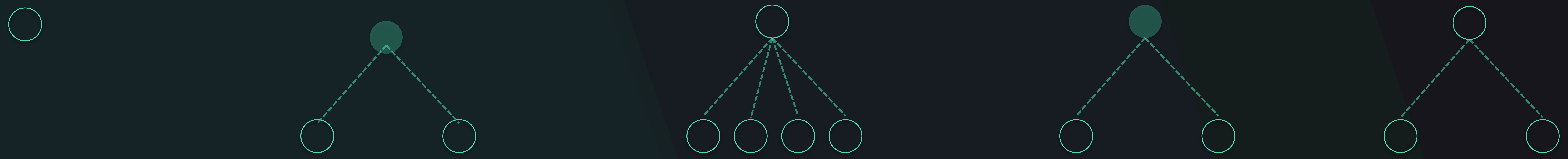


GRAPHISTRY

info@graphistry.com

No code – Graph makes AI interactive

Search | Connect Dots | Expand | Command | Drill



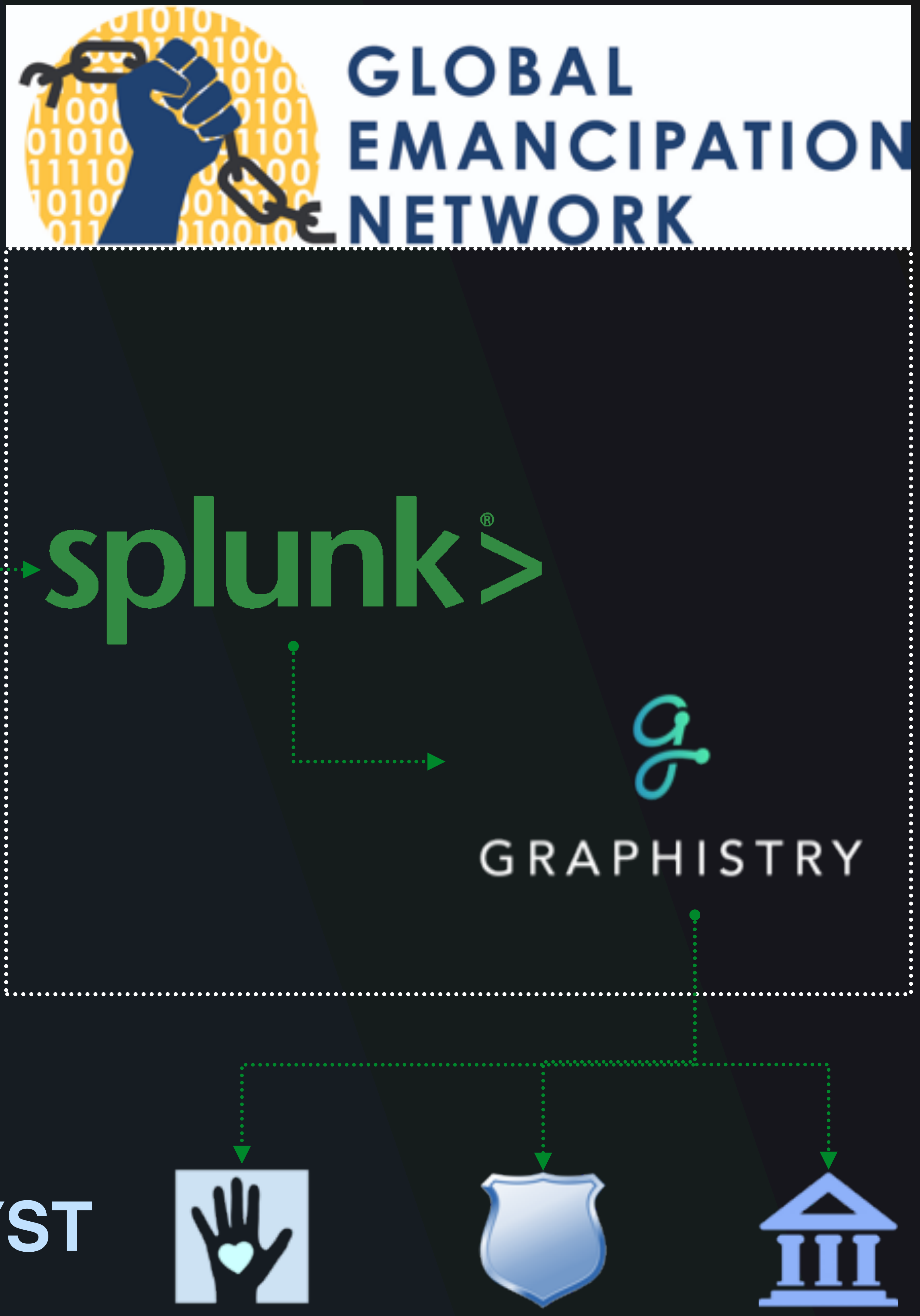
Work with
your DBs, APIs, enterprise, and algorithms
as one big *virtual* graph

Demo: Assisting the fight against human trafficking

DEMO: Find missing persons over anonymized digital trails
Generalize: Understand individual journeys

DEMO: Reveal large organized criminal behavior
Generalize: Surface big fraud, crime, influence, …

DARK WEB FORUMS

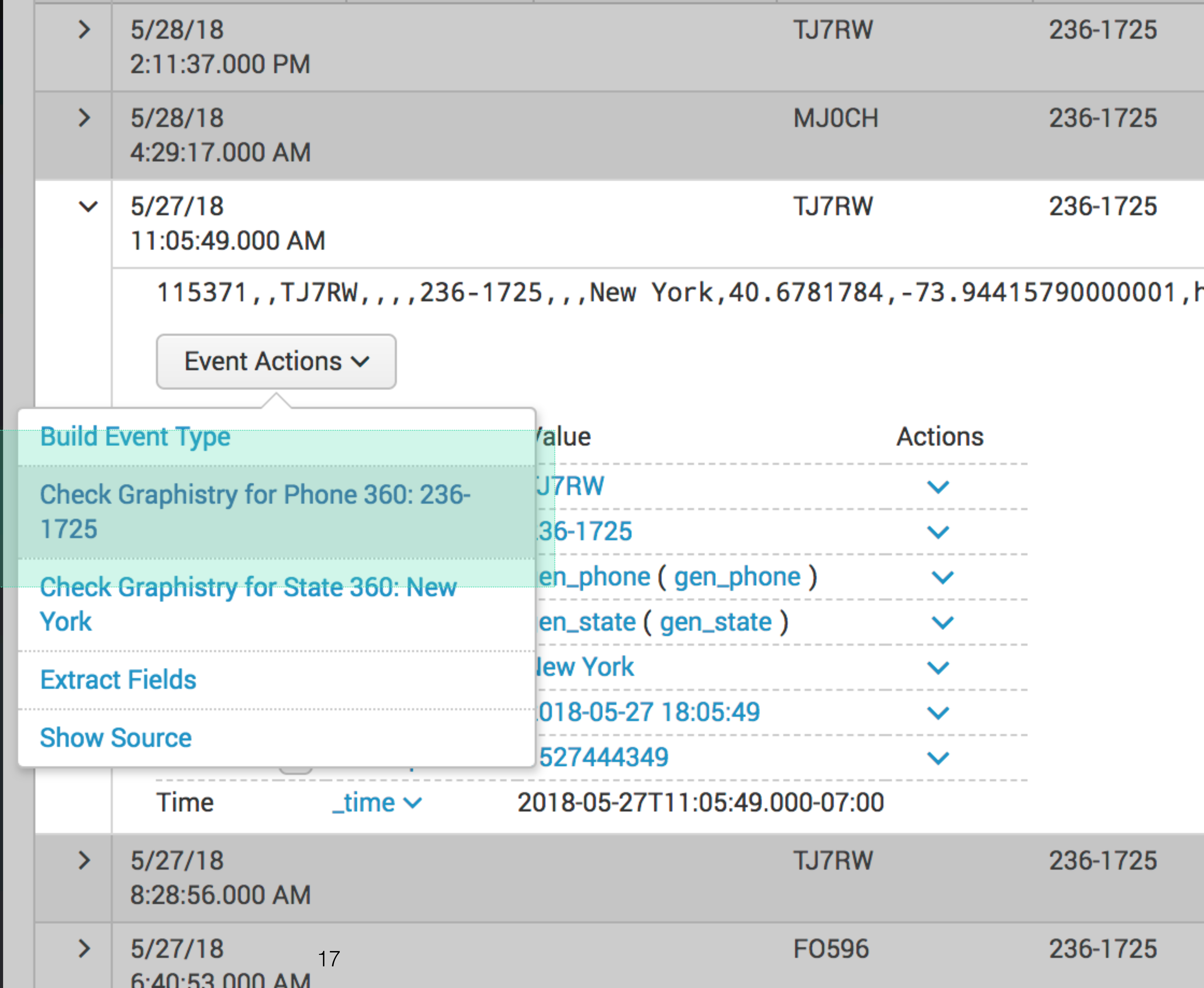


Data: Forum posts with shifting digital fingerprints -- names, emails, social network accounts, phones, locations, IPs, …

ANALYST

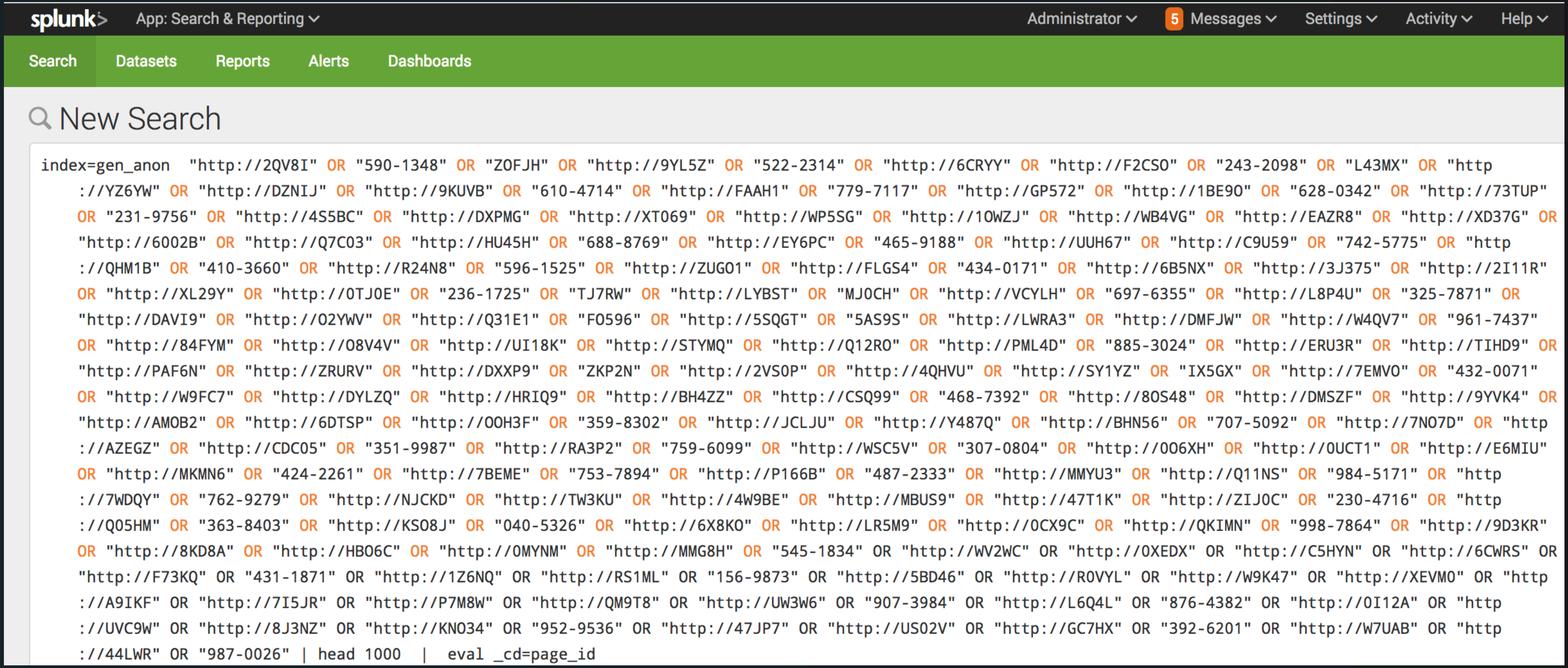
Splunk tables as a Virtual Graph

1. Easy expand
2. Easy automation

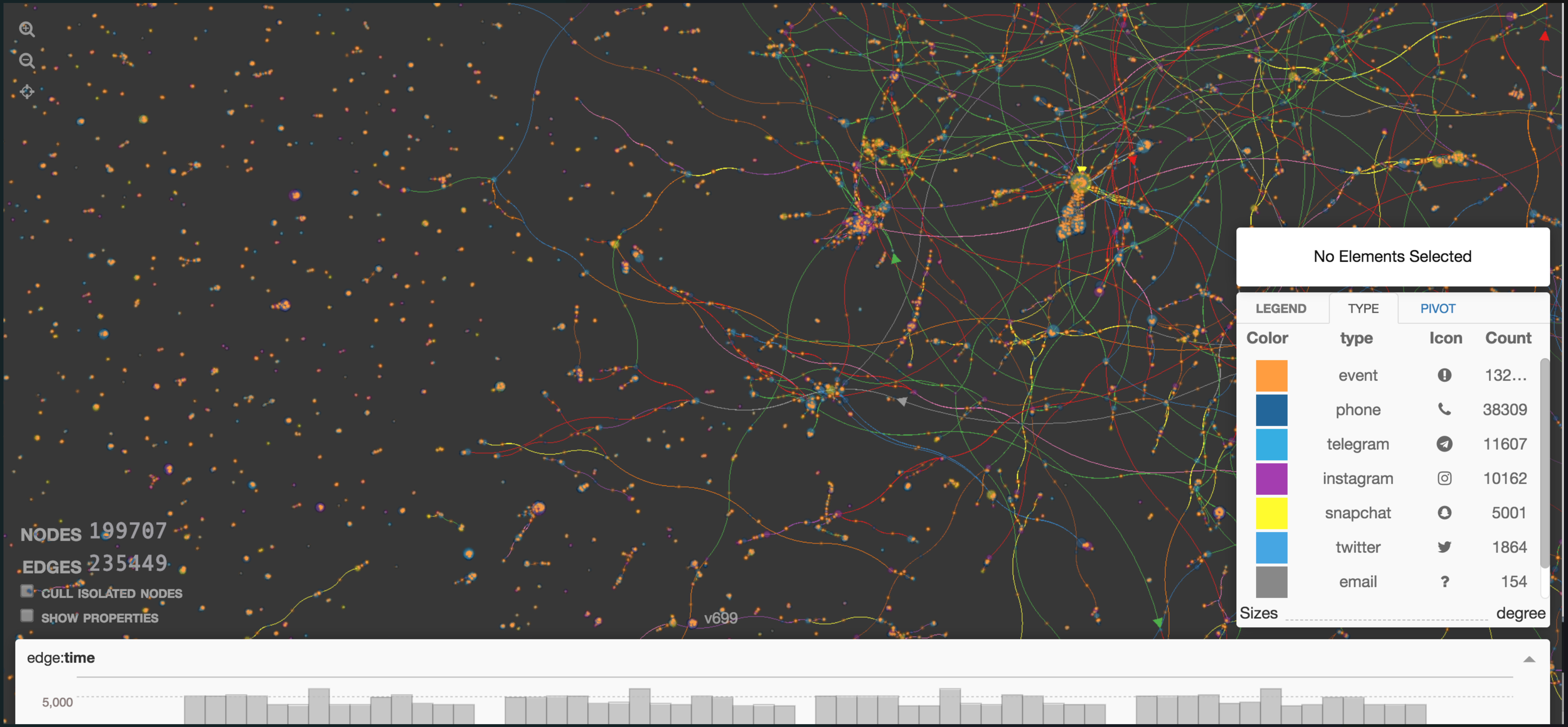


Auto-track across rotating IPs, browsers, phones, social media, ...

Generated data gathering query (#3 of 4)



Surface organized behavior by loading *all* flagged incidents



GRAPHISTRY

info@graphistry.com

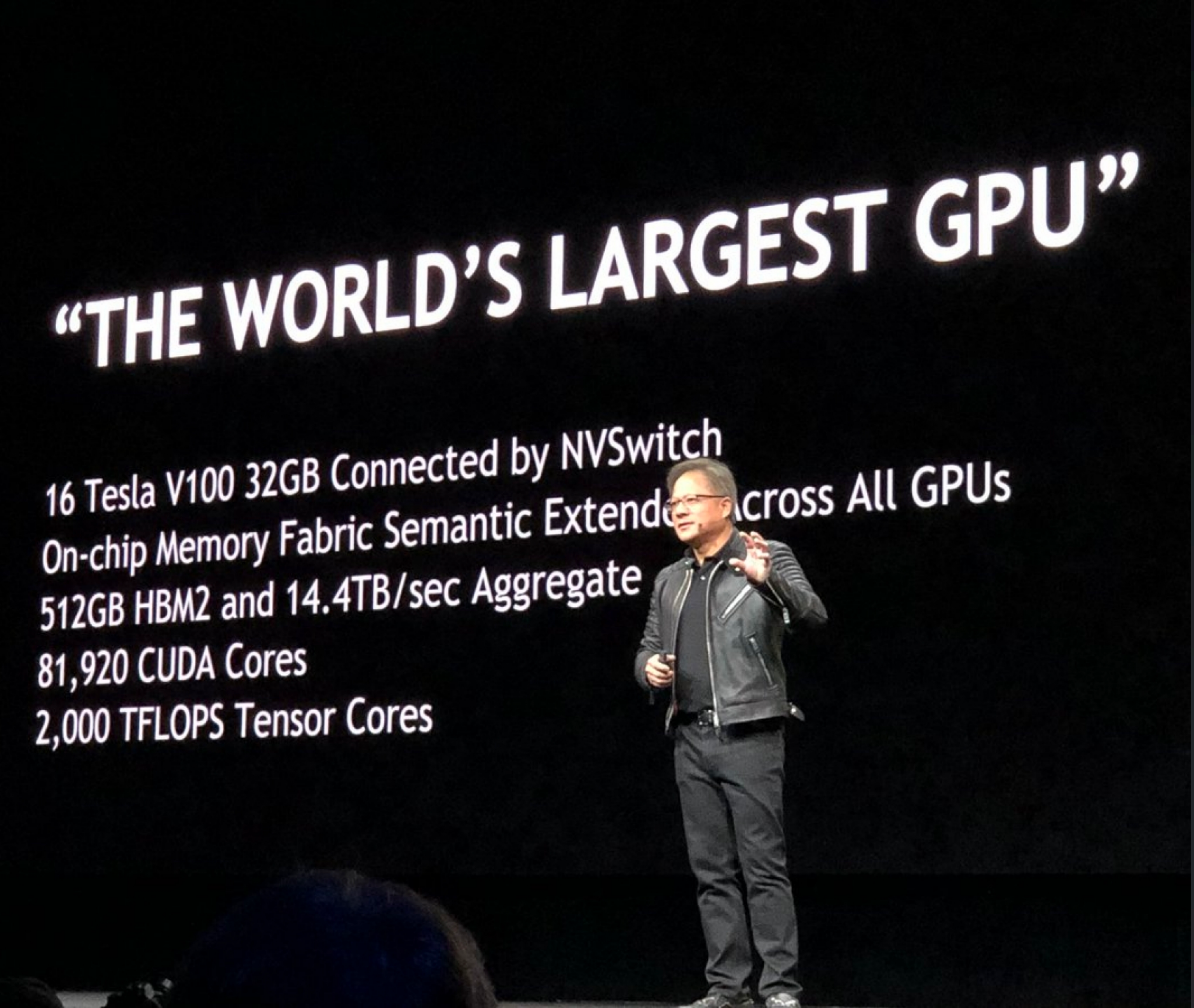
WHY BIG GRAPHS: Streetlight effect for digital era

Digital trails require scale

- Many events x metadata
- … per every entity!

Rapid hypothesis iteration

- Graphs move the spotlight
- GPUs enlarge the spotlight



The New Normal: Unified GPU client/cloud

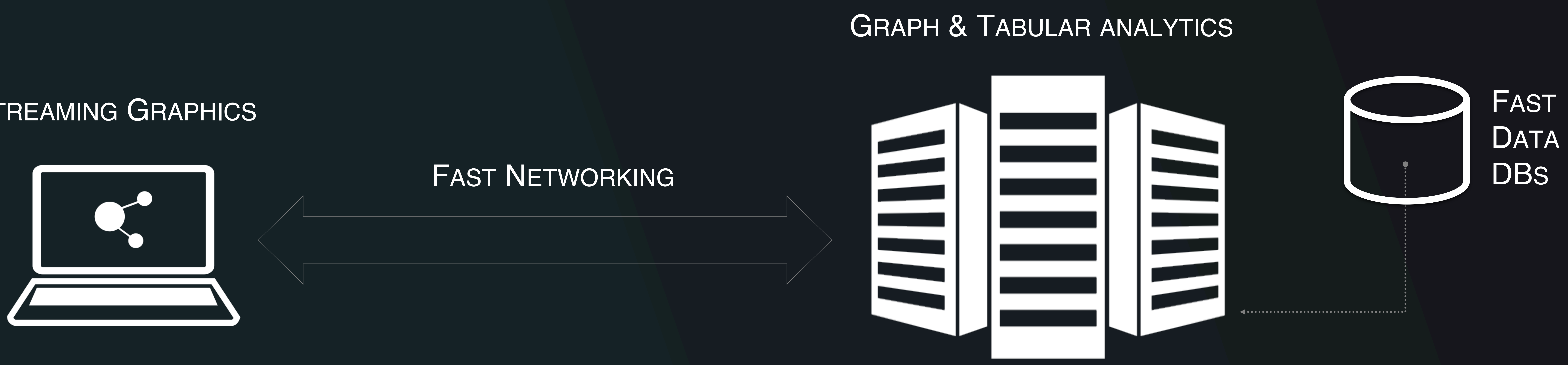
- GPU clouds
- Bandwidth can stream video
- Even phones have GPUs

Magic in 100ms

Graphistry: The first end-to-end GPU cloud visual analytics platform

GRAPHISTRY

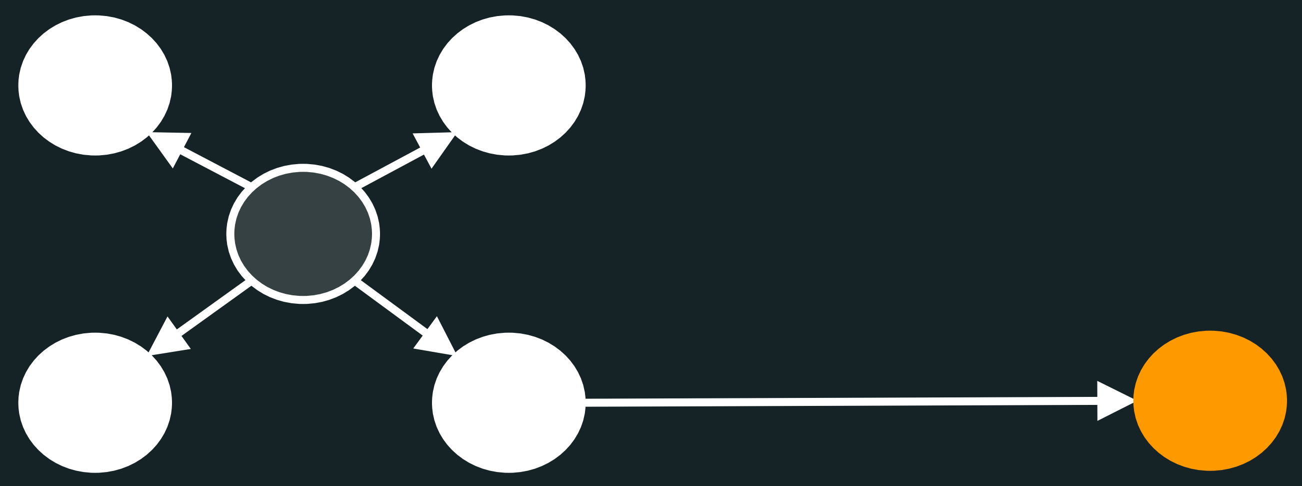
goai



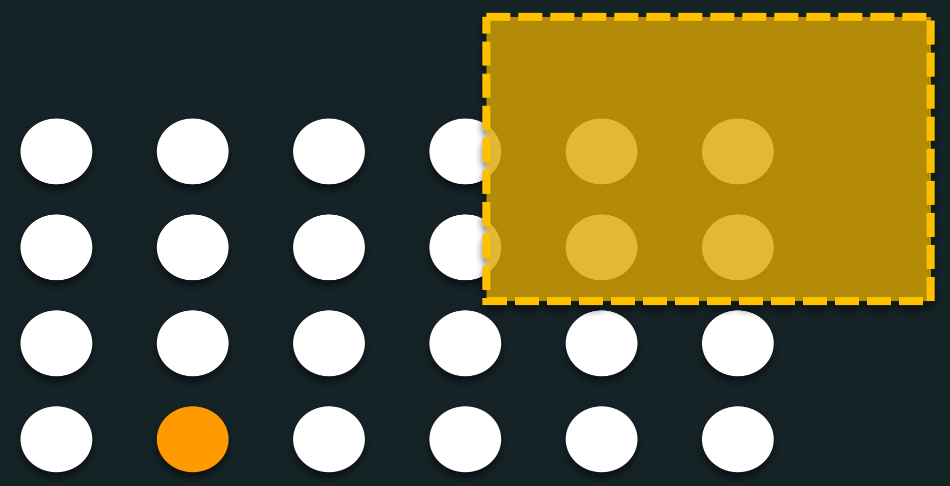
20ms ->10ms + <- 40ms 30ms 1-10s

GPUs: Scale Every Component 10-1000X in Speed + Data

Interactive Rendering
3+ million nodes+edges: 100X+ over D3.js



Meaningful Viz: Layout & ML
Smart clustering, coloring, sizing: 50X+ over Gephi



Interactive Analytics
1 NVidia Tesla K80 = ~9 TFLOPS

Problem: Jerky – 3fps

TIME

server		browser
broadcast	`points.buf` →	
ML tick		
ML tick		
ML tick		download
		render

info@graphistry.com

Problem: Jerky – 3fps

TIME

server

broadcast
ML tick
ML tick
ML tick
broadcast
ML tick
ML tick
ML tick
broadcast
...

`points.buf`

`points'.buf`

browser

download
render
download
render

2-3fps ☹

60fps Apache Arrow[JS] streams - 10KB-100KB chunks direct to GPU

TIME	server	browser		
	broadcast			
	ML tick	download.1		
	ML tick	download.2	render	smoooooth
	broadcast	download.3	render	
	ML tick	download.1	render	
	ML tick	download.2	render	interruptable
	broadcast	download.3	render	
	…	download.1	render	
		download.2	render	
		download.3	render	
			render	

GRAPHISTRY

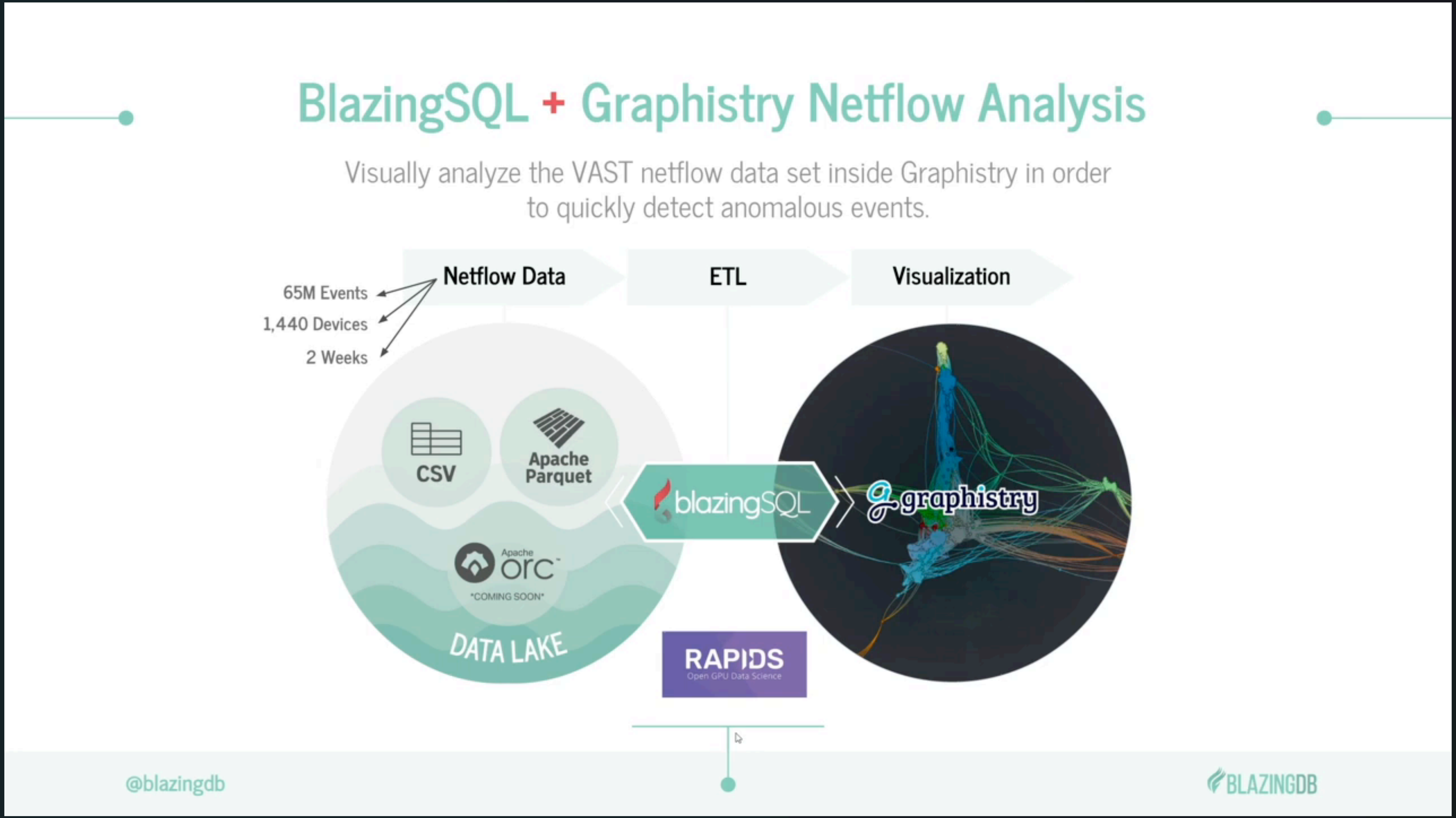
info@graphistry.com

Netflow – Working with data lakes with + BLAZINGDB

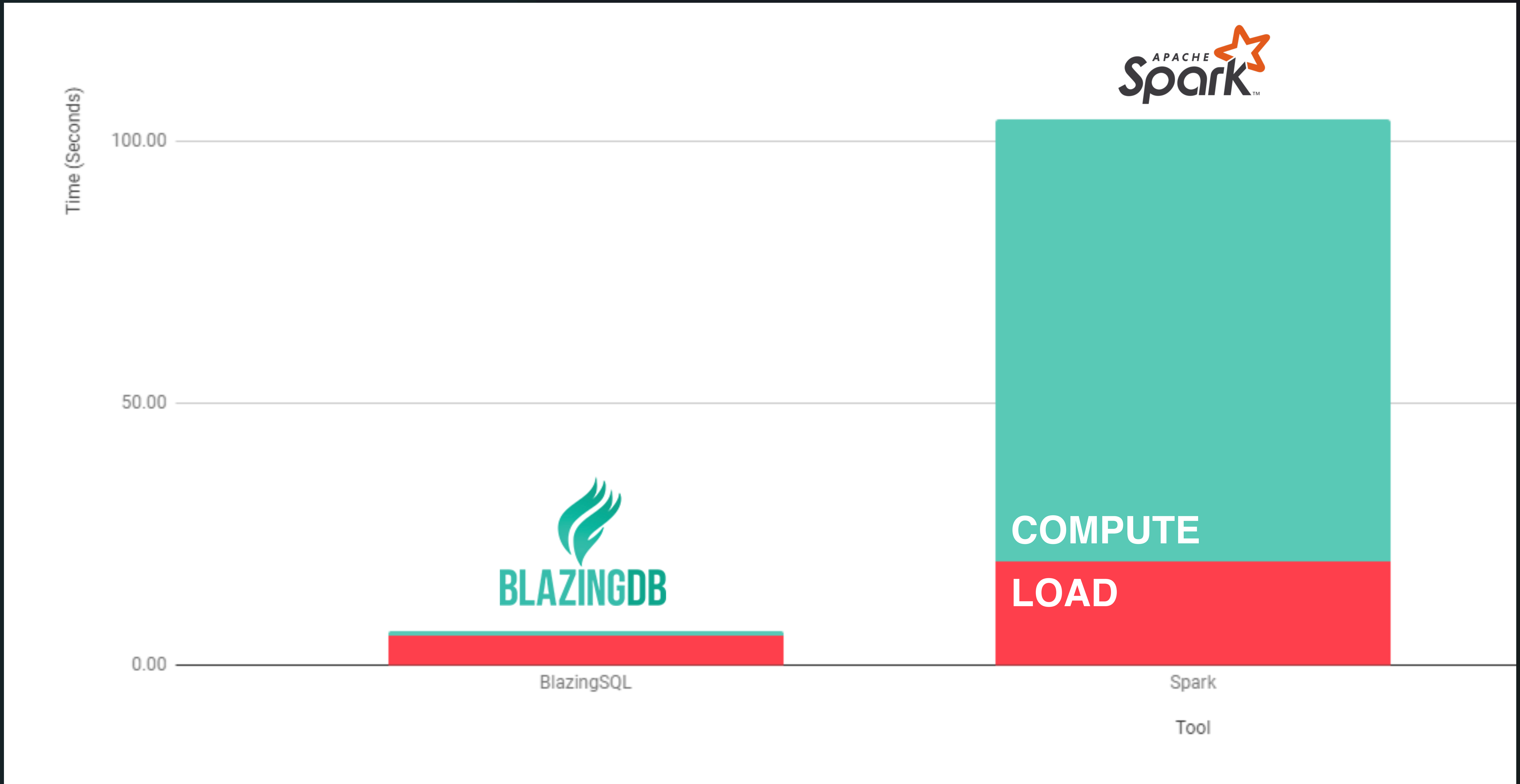
- 4M events / second
- Security, operations, …

- Goal: 300 node CPU cluster → 1 GPU node with point-and-click UI

Video


BlazingSQL + Graphistry Netflow Analysis
Visually analyze the VAST netflow data set inside Graphistry in order to quickly detect anomalous events.
Netflow Data
ETL
Visualization
65M Events
1,440 Devices
2 Weeks
CSV
Apache Parquet
Apache orc
COMING SOON
DATA LAKE
blazingSQL
graphistry
RAPIDS
Open GPU Data Science
@blazingdb
BLAZINGDB

For same $:   2s vs. 1.5min



EVENT ANALYTICS
fraud, threat intelligence, it security, influence, counter-trafficking, operations, …

Journey to 100X'ing Investigations

- Why big graphs
- No code for ML & automation
- GPUs: GOAI in Graphistry Engine 2.0

info@graphistry.com

- Graphistry Engine 2.0!
- Launching Private Cloud (NB/API users)
- Early Users: Workflow Automator (Splunk, ELK, Neoj, JanusGraph, SQL, Google Chronicle, MS Sentinel, …)