Working through many events with GPUs

BlueHat, Redmond, September 2018

GRAPHISTRY

Powering the future of human-in-the-loop investigations

Leo Meyerovich
CEO & Janitor
@lmeyerov

Who

- Team: Berkeley, Netflix, RedSeal, Fortinet, …
- Users: Cybersecurity, anti-fraud, AML, digital crime, …

  Researchers, developers, & analysts

What

Powering the future of visual event analysis:

✓ Understand: Visual analytics w/ GPUs & Graphs

✓ Workflow: Visual investigation templates, APIs, …

GRAPHISTRY

info@graphistry.com

Today

Working with big graphs

- **Why: logs logs logs logs logs**
- **Visualization & GPUs in 2018 & 2020**
- **Structured programming (templates) for visual investigations**

Applied to Digital Crime

- Trafficking & Theft
- Security Outbreaks, Malware Infections, & Network Hunting

Demo

- Ethereum theft: $1M / week
- Quickly map out large set of 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

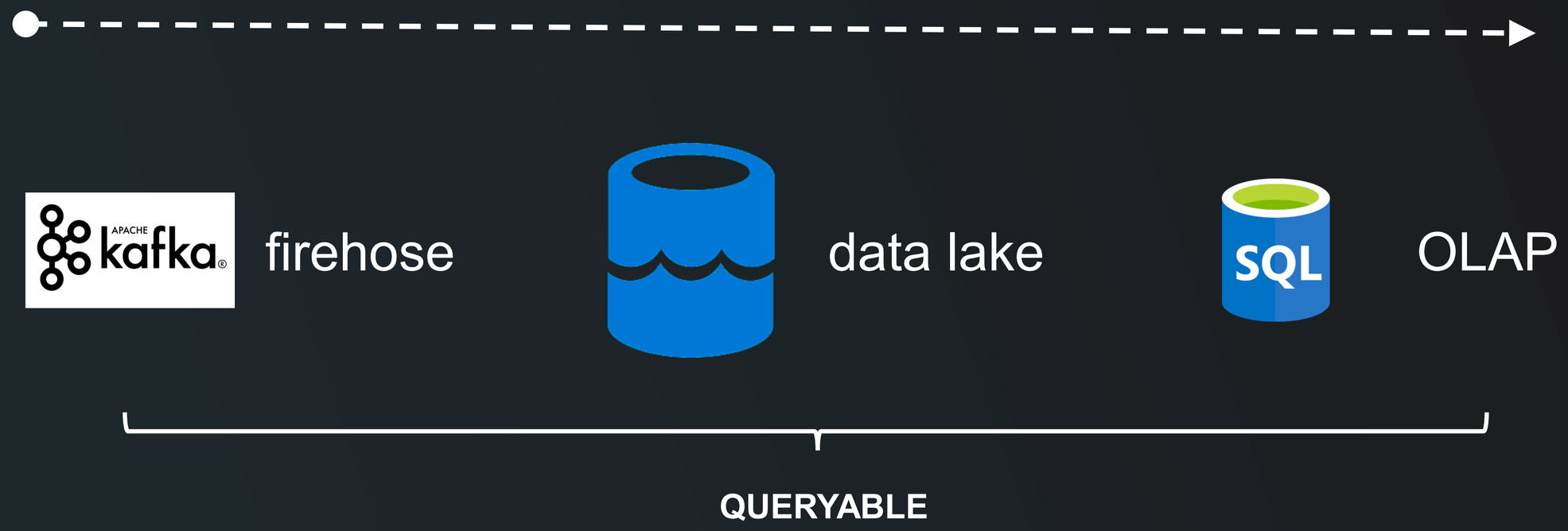
2018-08-29

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。

Enterprise Graph: Logs rules everything around me

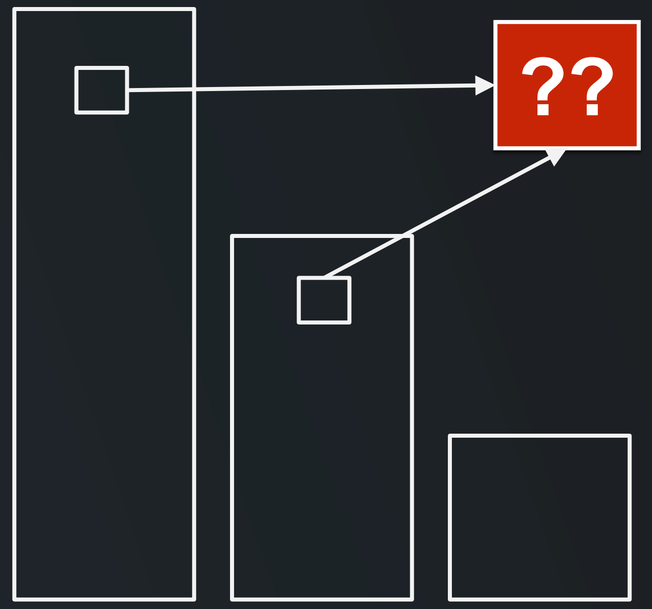
Events all the way down

e1(ip1, mac1, alertX, ip2, mac2, …)
e2(ip2, mac2, alertX, ip1, mac1, …)



Bar charts, heatmaps, …: Low-Dimensional & opaque

```
> $  select * from ALERTS where PRIORITY > 8
```

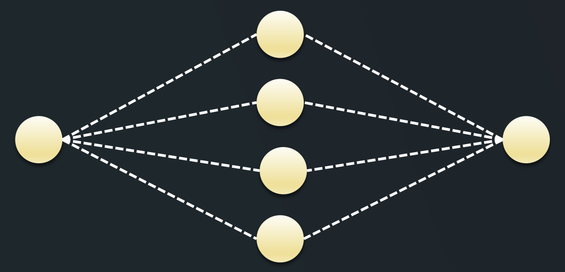


Demo: Trafficking - GlobalEmancipationNetwork.ngo

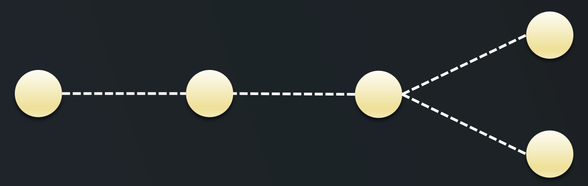
- Data: Dark web scraping
- Task: Missing persons & Interstate crime networks
- Ex: Given a phone number, what can we see?

Graph: Lingua Franca for Undertanding & Interacting with Data

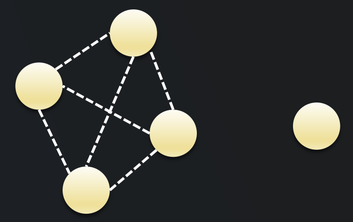
Events, entities, & scope



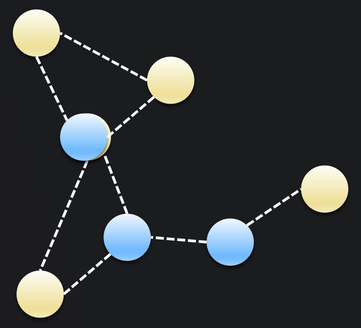
Progression & behavior



Correlations & outliers

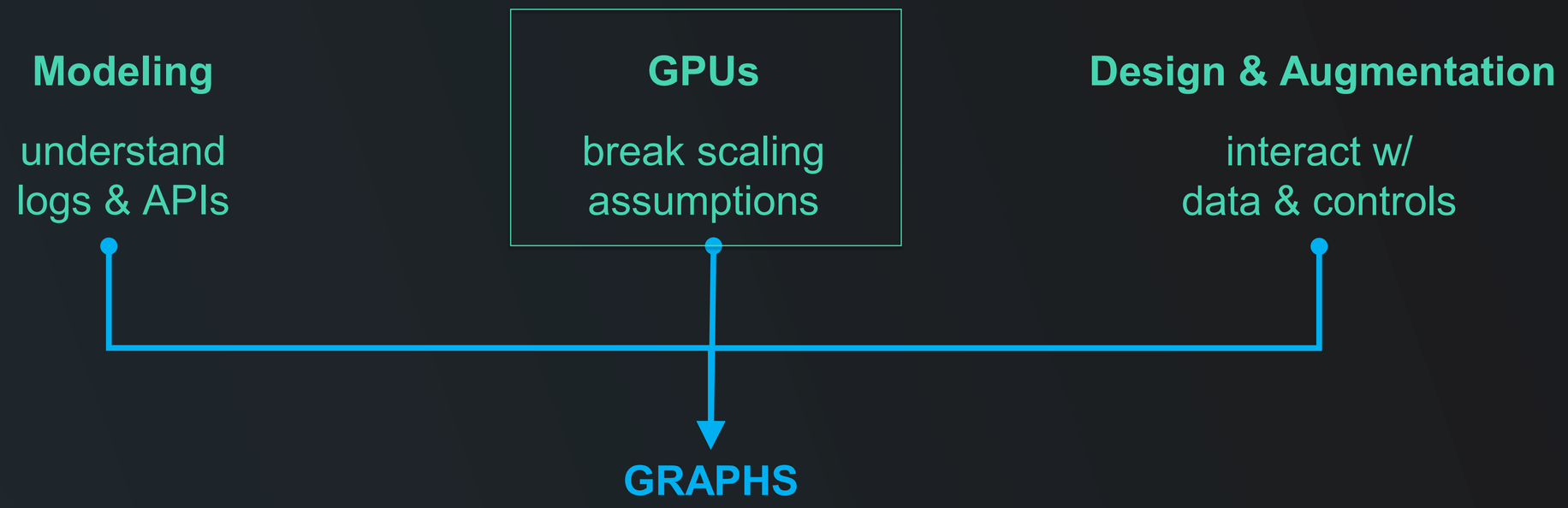


Influence



Interact: Expand, drill, find paths, cluster, …

100Xing Human-in-the-Loop

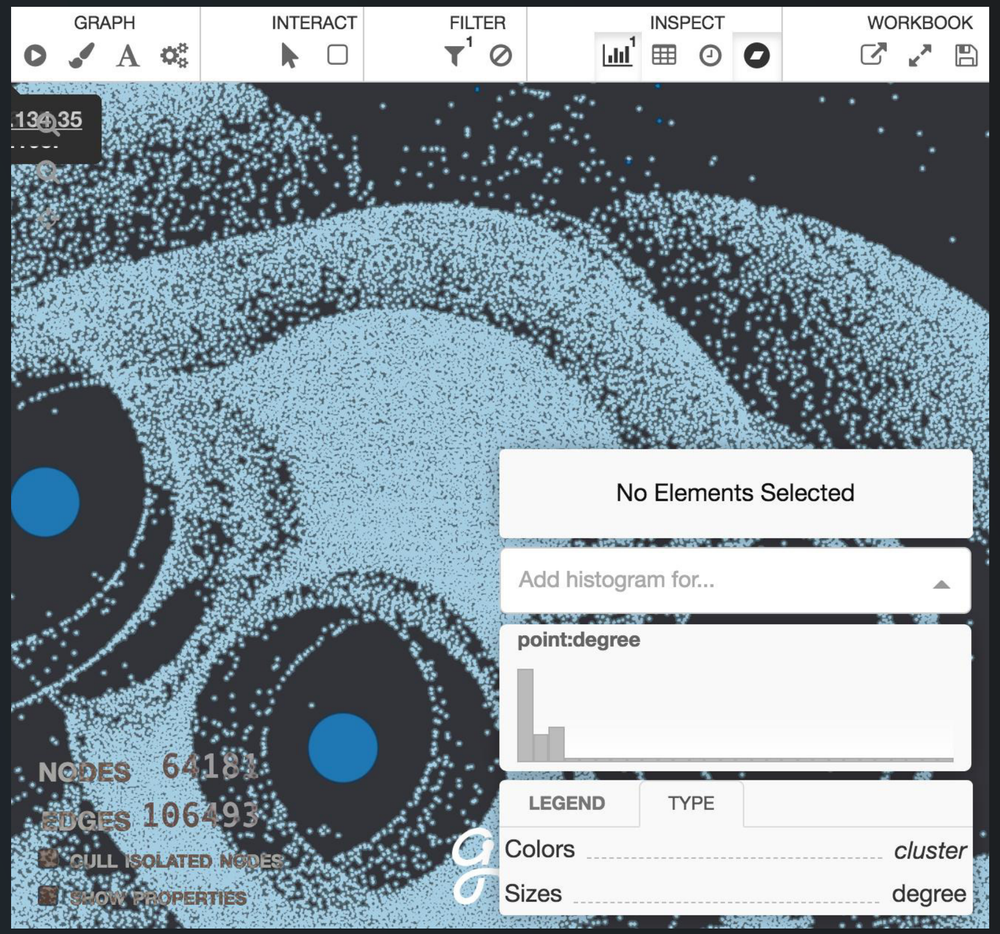


WHY BIG GRAPHS: Streetlight Effect

Rapid hypothesis iteration

- Graphs move the spotlight
- GPUs enlarges the spotlight

WHY GRAPHS: Situational Awareness Demo



Graphistry Users Do the Darndest Things <3 <3 <3

- Digital transactions (blockchain, …)
- Machine & network logs
- Situational awareness: Assets, IoT, outbreaks, …
- Territory analysis: trafficking, …
- OSINT: Political influence, Twitter scams, botnets & DNS, …

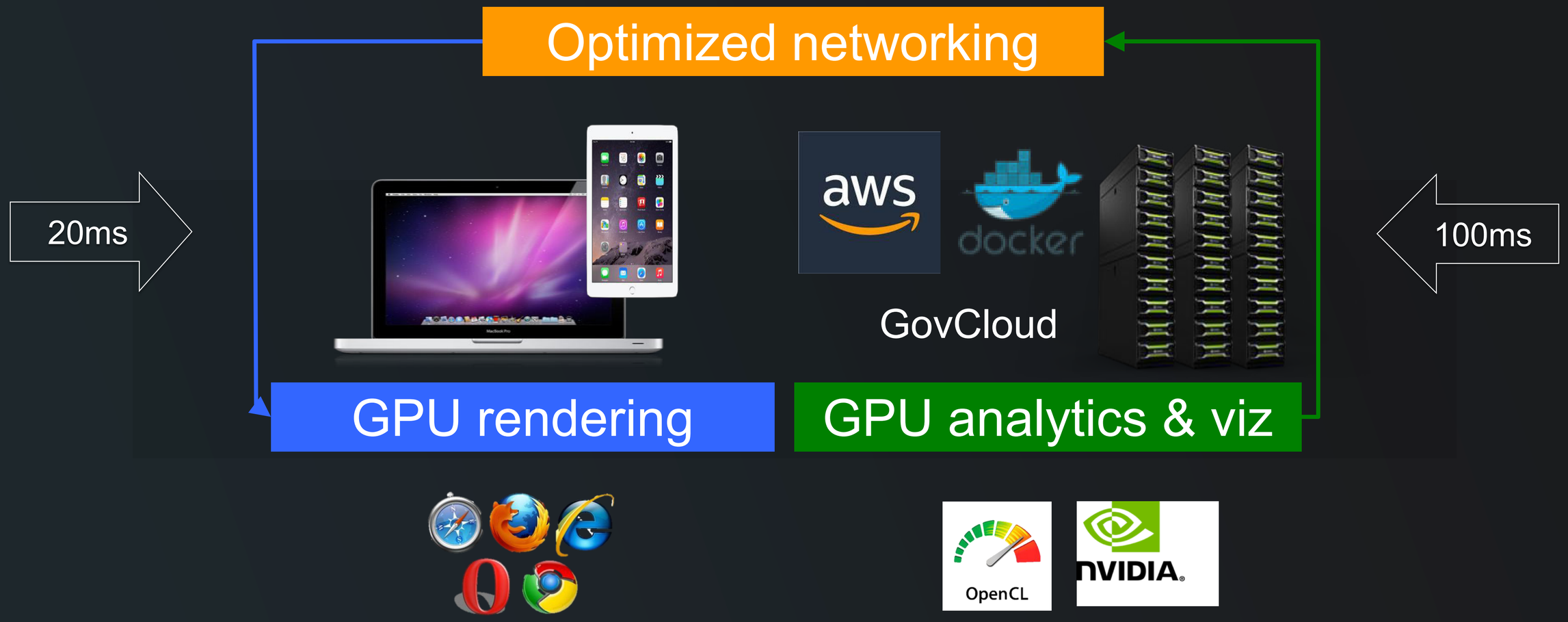
Achieving perceptual magic with GPUs

RESPONSIVE

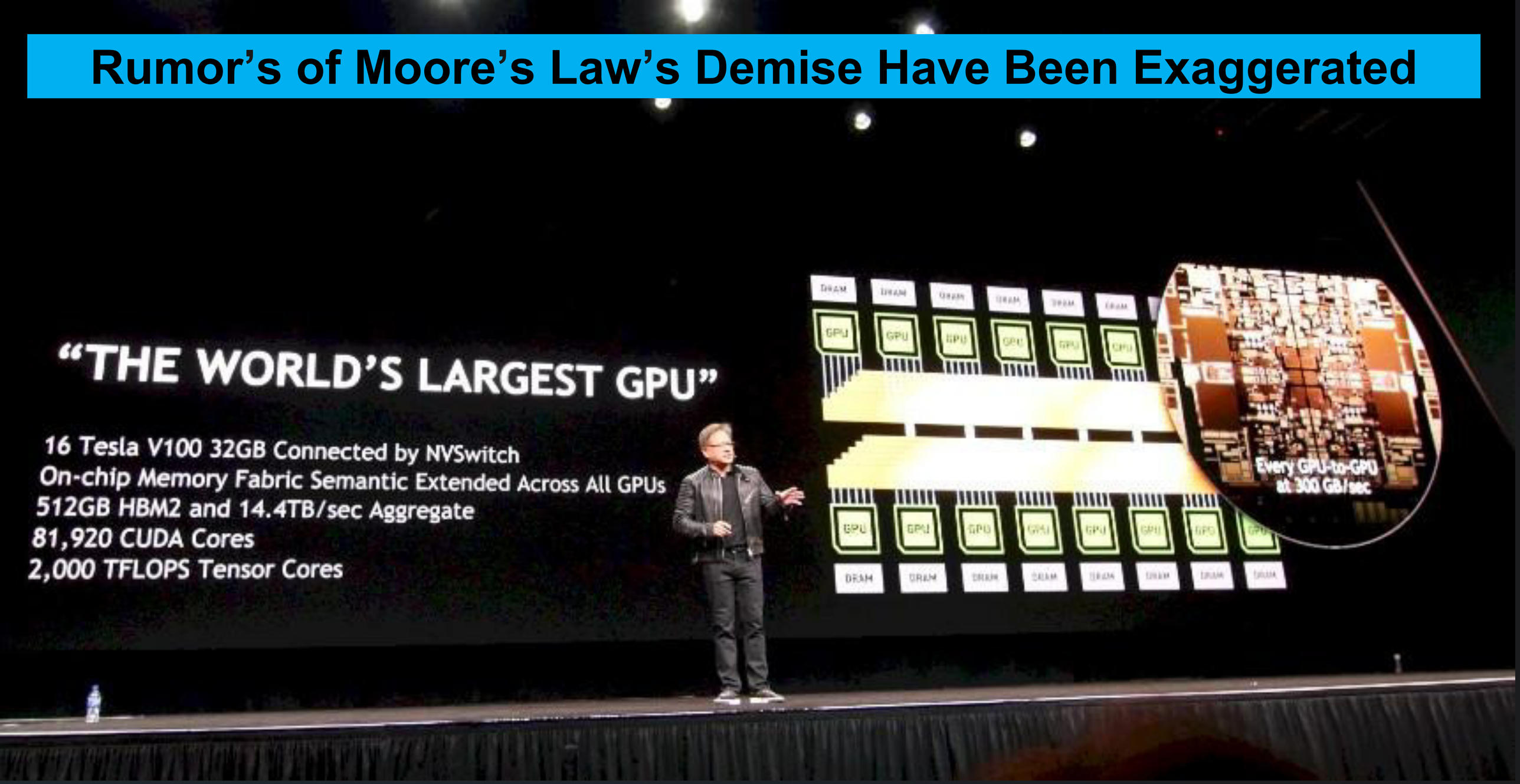
5ms	inner ear
16ms	60fps pan/zoom
100ms	clicking things
1-3s	typing stuff
> 3s	legacy

info@graphistry.com

Graphistry Redefines Your Device: GPU Client <> GPU Cloud



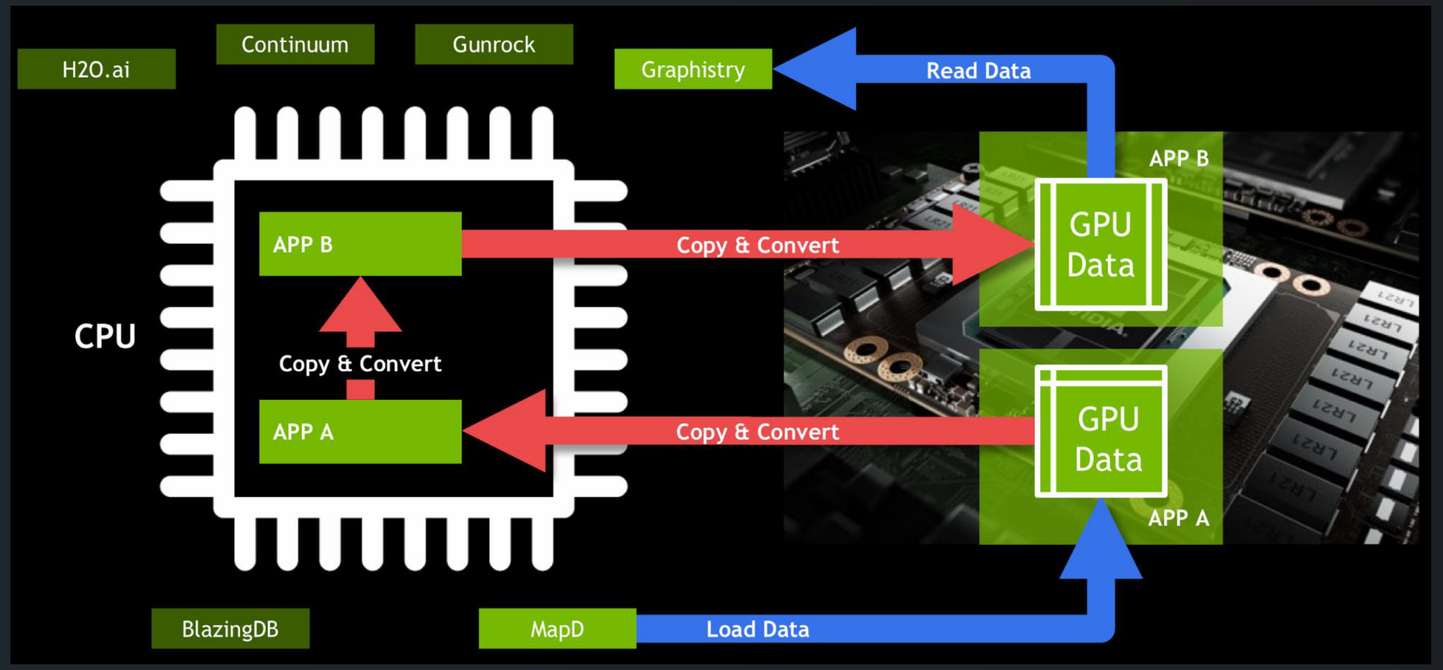
Rumor's of Moore's Law's Demise Have Been Exaggerated



Hadoop moment for GPU Fast Data: 1B rows / 100ms

GOAi GPU OPEN ANALYTICS INITIATIVE

GPU Open Analytics Initiative

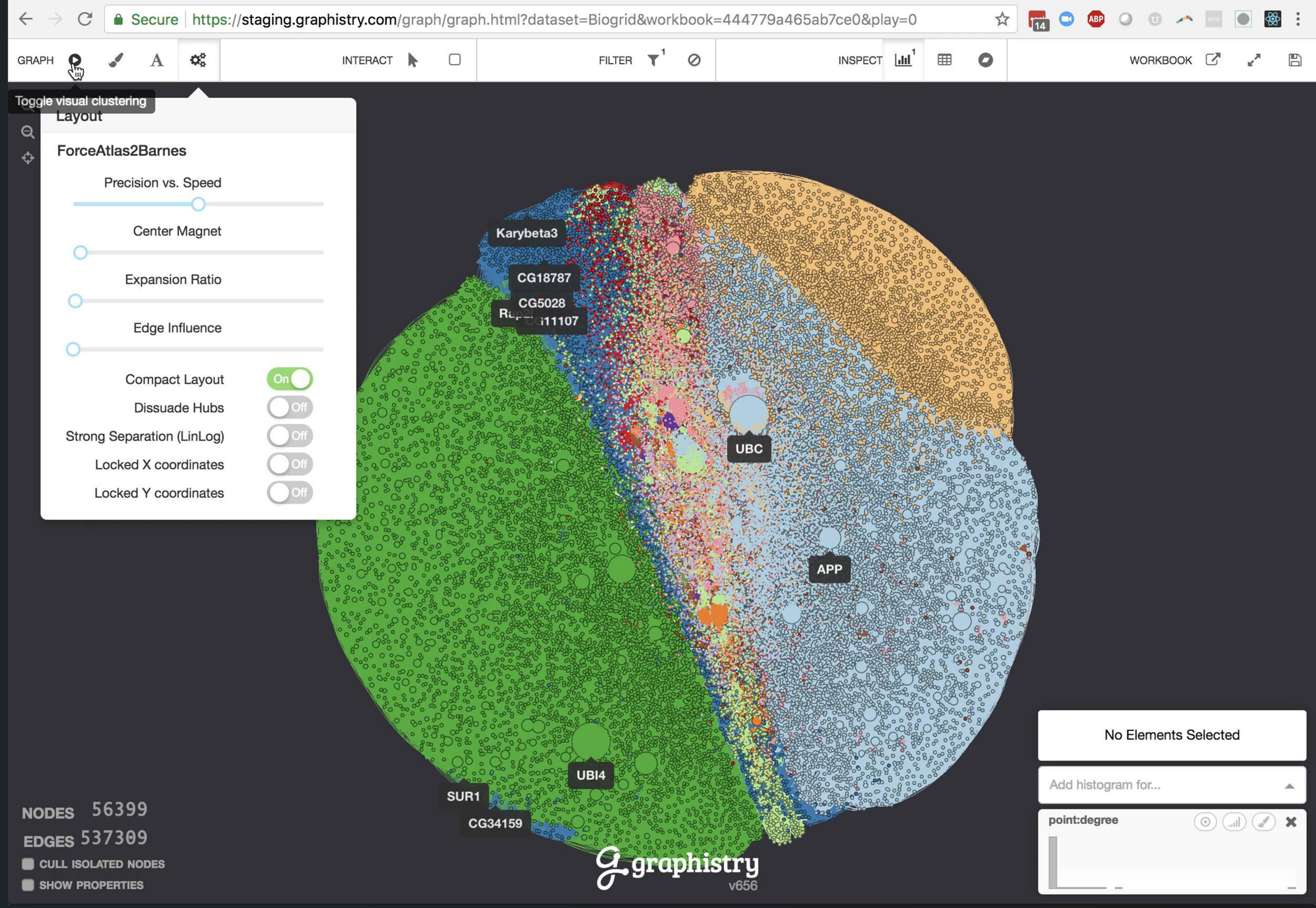


THE APACHE™ SOFTWARE FOUNDATION Arrow

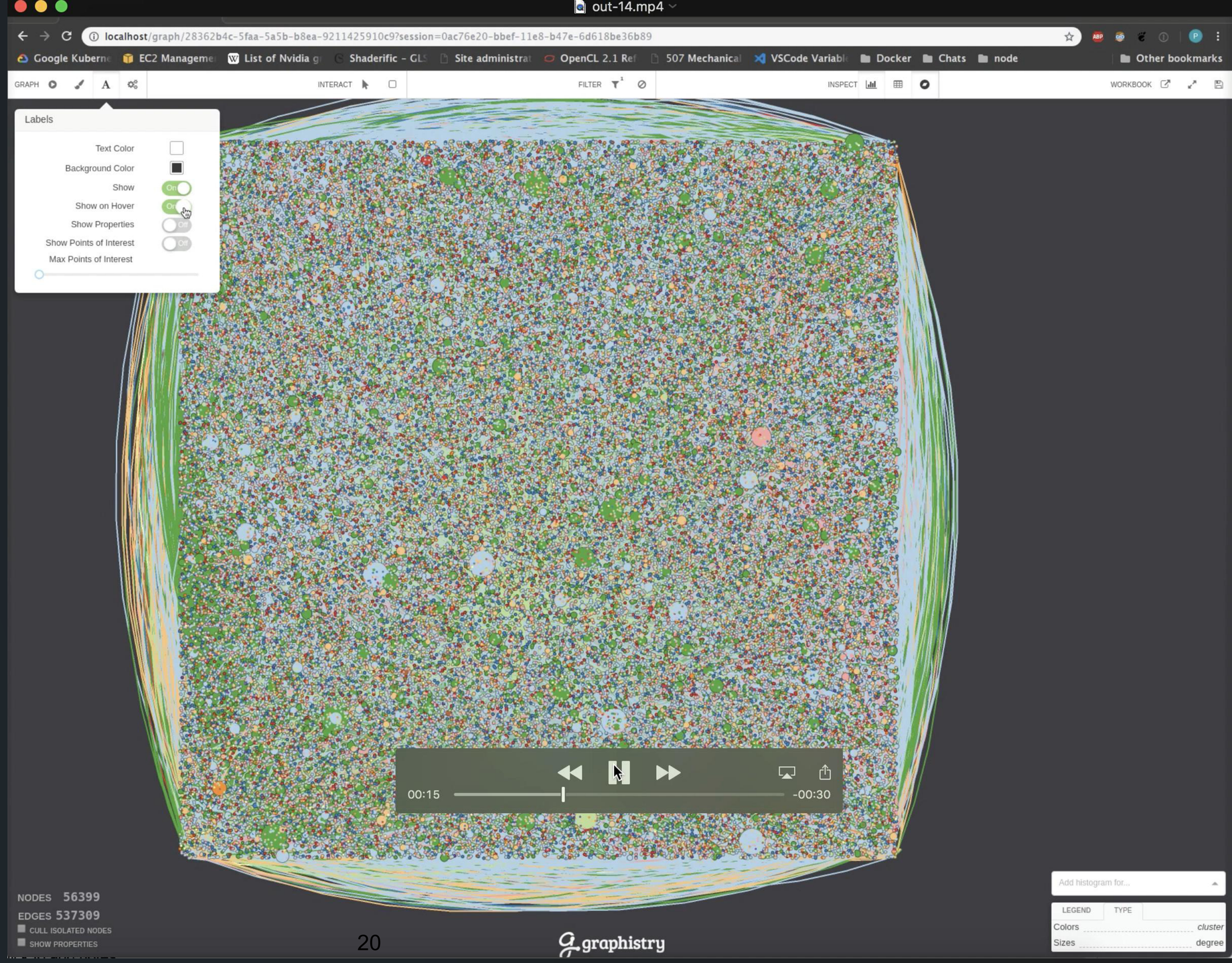
- Numba: Python -> CUDA
- PyGDF: GPU dataframes
- NVGraph: GPU graph

info@graphistry.com

Demo: Graphistry 2.0

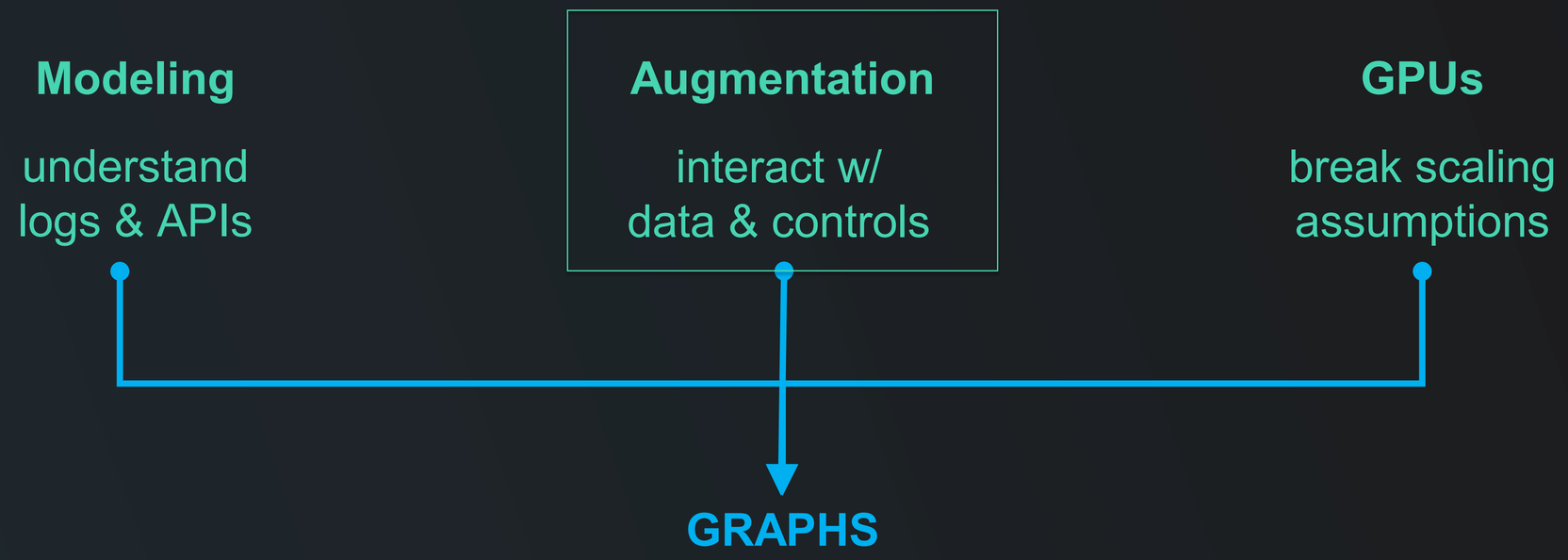


Demo: Graphistry 2.0



GRAPHISTRY

100Xing Human-in-the-Loop



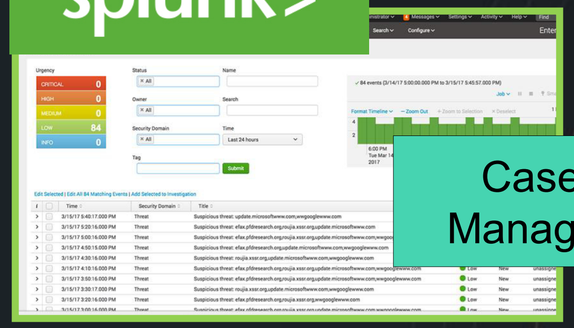
Problem:
Graph investigation TAKES TOO MUCH TIME

1-3 min	• Find/query data
0-10 min	• Clean
2-10 min	• Analyze
	• … repeat
	➔ Act

Template Case Study: Augmenting splunk> incident handling

GRAPHISTRY

splunk>



1. Open incident

Case Manager

Prebuilt Template

splunk>

FireEye

CYLANCE

paloalto NETWORKS

2. Auto-gather & correlate data

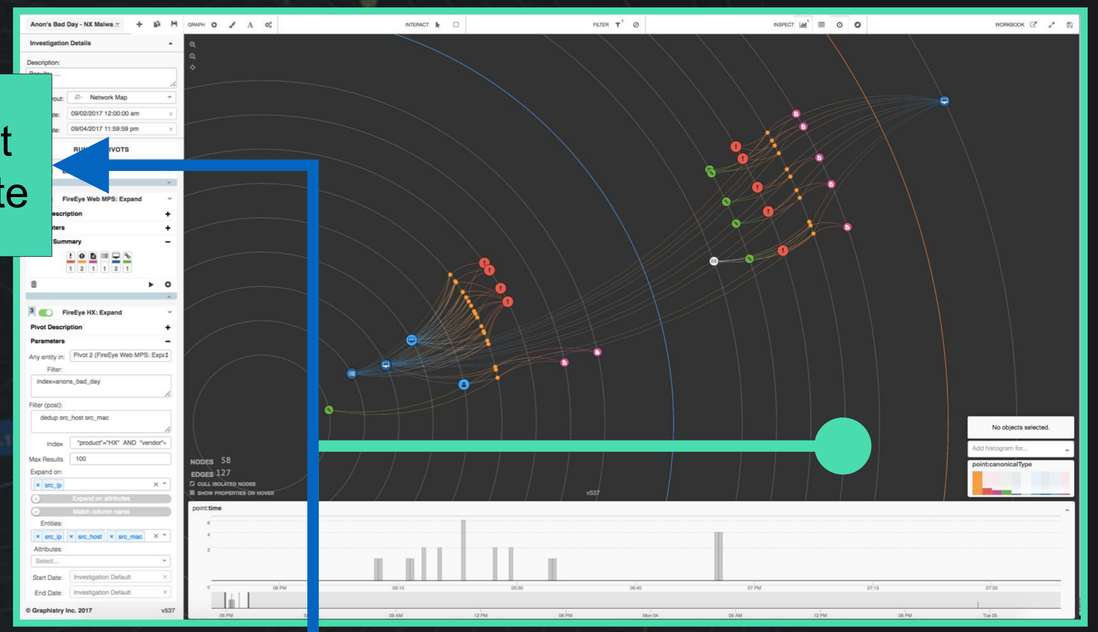
Databases & APIs

Spark

...

VirusTotal

goai



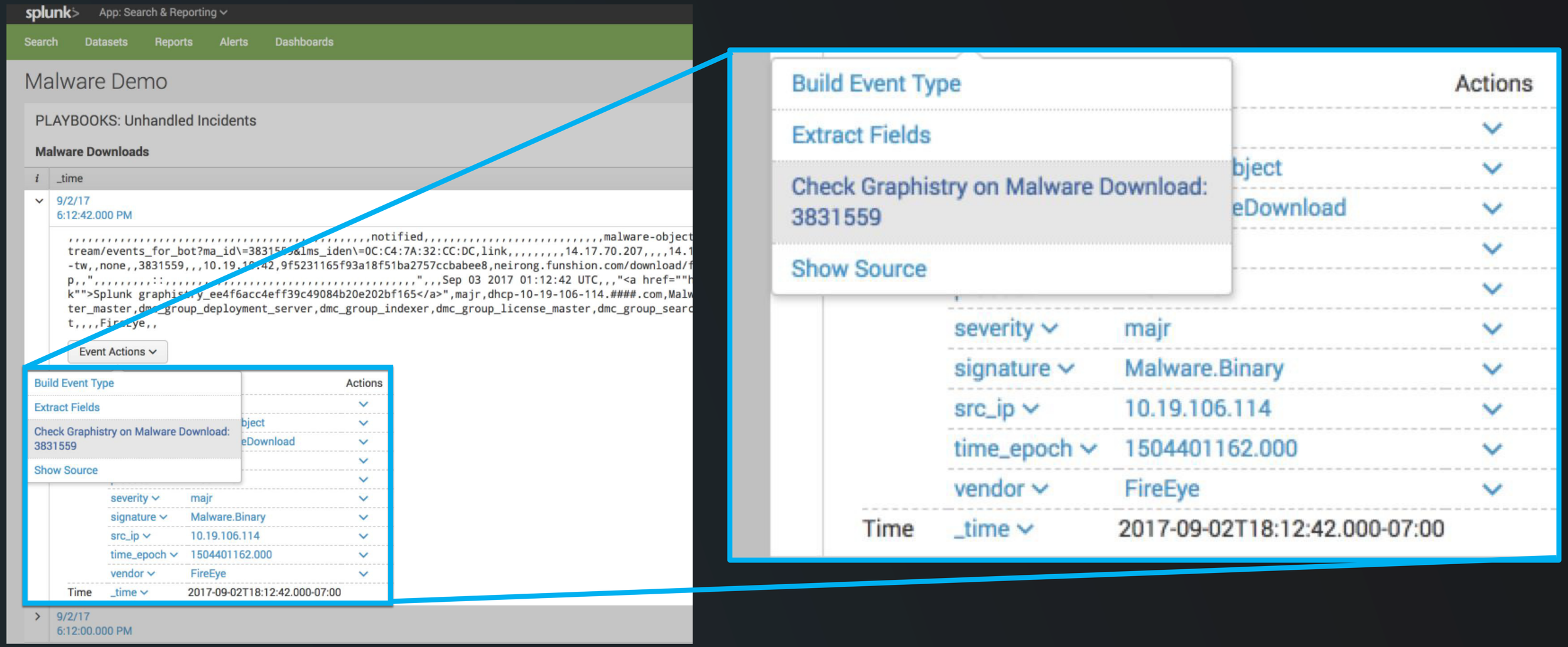
1 server in cloud or on-premises

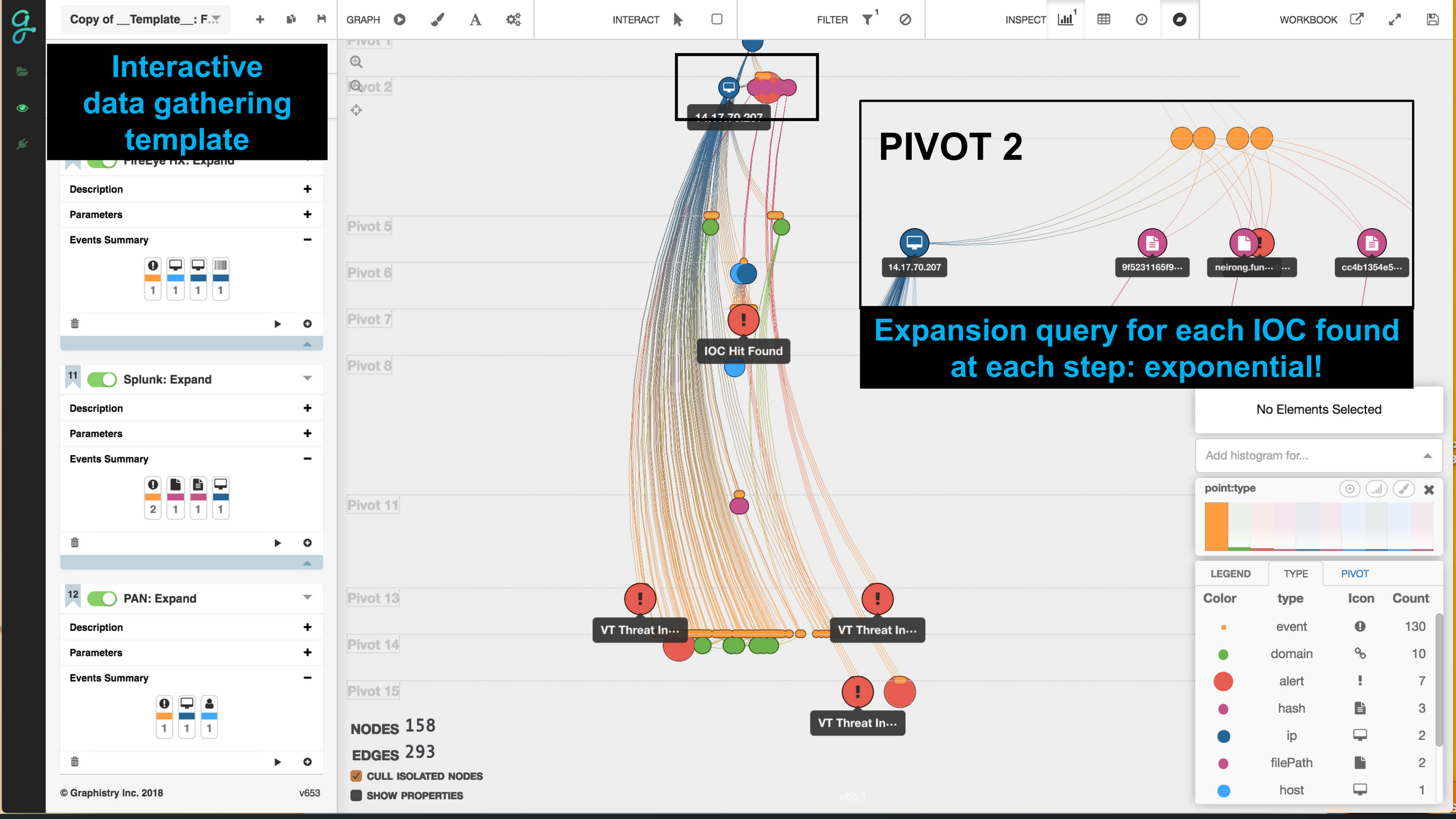
GPU-Accelerated Visual Graph Analytics

NVIDIA

3. See Explanation & Visually Drill

DEMO: Workflow (Malware)




Copy of __Template__: F...
GRAPH
INTERACT
FILTER
INSPECT
WORKBOOK
Interactive data gathering template
Description
Parameters
Events Summary
11 Splunk: Expand
12 PAN: Expand
© Graphistry Inc. 2018
v653
Pivot 2
Pivot 5
Pivot 6
Pivot 7
Pivot 8
Pivot 11
Pivot 13
Pivot 14
Pivot 15
14.17.70.207
IOC Hit Found
VT Threat In...
NODES 158
EDGES 293
CULL ISOLATED NODES
SHOW PROPERTIES
PIVOT 2
14.17.70.207
9f5231165f9...
neirong.fun...
cc4b1354e5...
Expansion query for each IOC found at each step: exponential!
No Elements Selected
Add histogram for...
point:type
LEGEND
TYPE
PIVOT
Color
type
Icon
Count
event 130
domain 10
alert 7
hash 3
ip 2
filePath 2
host 1

GRAPH INTERACT FILTER INSPECT WORKBOOK

Internal Network

IOC Hit Found

21614D1F38...

VT Threat In...

9f5231165f9...

cc4b1354e5...

malware-ob...

14.17.70.207

NODES 12

EDGES 60

CULL ISOLATED NODES

SHOW PROPERTIES

No Elements Selected

Add histogram for...

point:type

LEGEND TYPE PIVOT

Color	type	Icon	Count
	alert	!	7
	file		3
	host		1
	domain		1

edge:time

10

5

0

06:10 06:15 06:20 06:25 06:30 06:35 06:40 06:45 06:50 06:55 07 PM

SAT SEP 02 2017 18:07:30 GMT-0700

BAR WIDTH: 30 SECONDS

SAT SEP 02 2017 19:02:30 GMT-0700

Structured Programming for Visual Investigations

- **Speed**: MTTR < 10min?
- **Insight**: More columns, more data sources, … more queries
- **Reliability**: Automate best practices for breadth & depth
- **Centralized Record**: Train analysts, manage sensor/control churn, …
- **Process**: Over time, built up library, harden flows, & increase coverage

Summary

Big graphs are here and we can handle

- Why: High dimensional correlation
- Scale perf 100X+ with GPUs for visibility & iteration
- Structured programming for visual investigations

➔ Augmented intelligence

GRAPHISTRY

@lmeyerov
info@graphistry

Non-sensitive data API keys:
graphistry.com

Data Future: Going from 1B to 1T+

Data loading	1T+ in 0ms with Arrow pointers
Many columns	400+ columns with graph
Many rows	1B/100ms: PyGDF single GPU 10B/100ms: PyGDF multi-GPU? 1T/100ms: PyGDF rack?
Client	GPU<>GPU arrow streaming

The Next 100X of Scale, Flexibility, & Intelligence

In active development; targeting end of year

goai

Apache Arrow
Faster conclusions using in-memory columnar SQL and machine learning

NVIDIA

Graph in user system, esp. Hadoop + Data Science ecosystem

nodes.arrow

edges.arrow

Apache Arrow:
In-memory columnar *dataframe* binary format

1. Ingest More, Faster

2. More GPU Analytics

Graphistry GPU Server

Graphistry GPU Graph Kernels

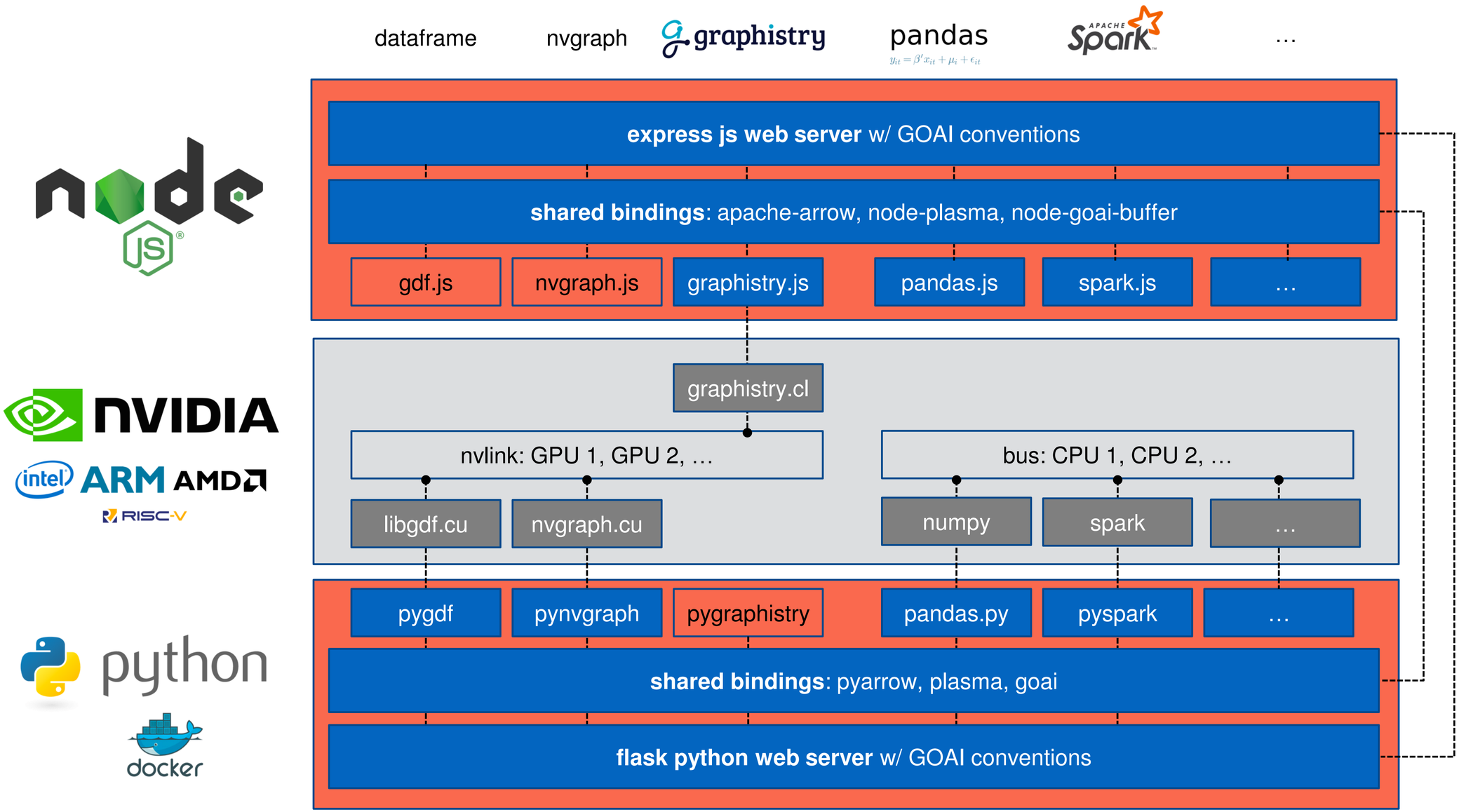
DB, API, …

nvGraph

Accelerated networking

Graphistry GPU Client

Graphistry GPU rendering engine, …

3. Developers: Bring GPU to Web via node, browser libs

Reference Architectures for JavaScript GOAI Bridge to GPU Everything