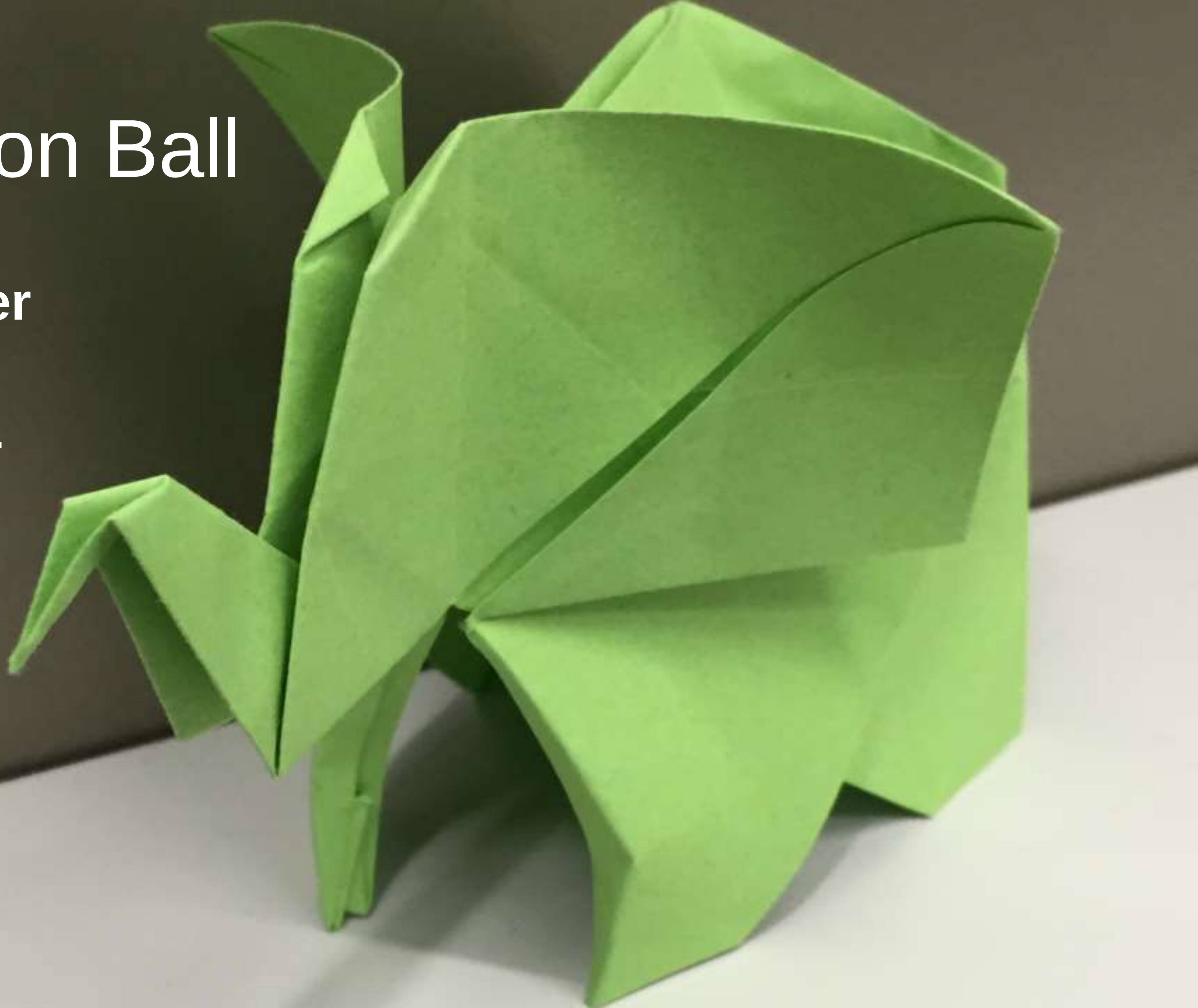
Solving Cyber at Scale

with Hadoop, Storm and Metron

Simon Elliston Ball

- **Product Manager**
- **Data Scientist**
- **Elephant herder**

🐦 **@sireb**



Threat Sources

IoT: Mirai

Reports of 1.2 Tbps
500,000 devices at peak
DDoS attacks on Dyn DNS services

Insiders

Ransomware and spears

Who are we up against?

MEECES

Money
Ego
Entertainment
Cause
Entrance (social acceptance)
Status

Big Business

- $tn market
- Access is bought and sold: 5 bitcoin for 100m accounts
- Sharing networks
- Criminals as a Service
- DDoS attacks: cost attackers $5 per hour, defenders ~$40k

Sources: BT and KPMG Report, Taking the Offensive

Challenges for the Modern SOC

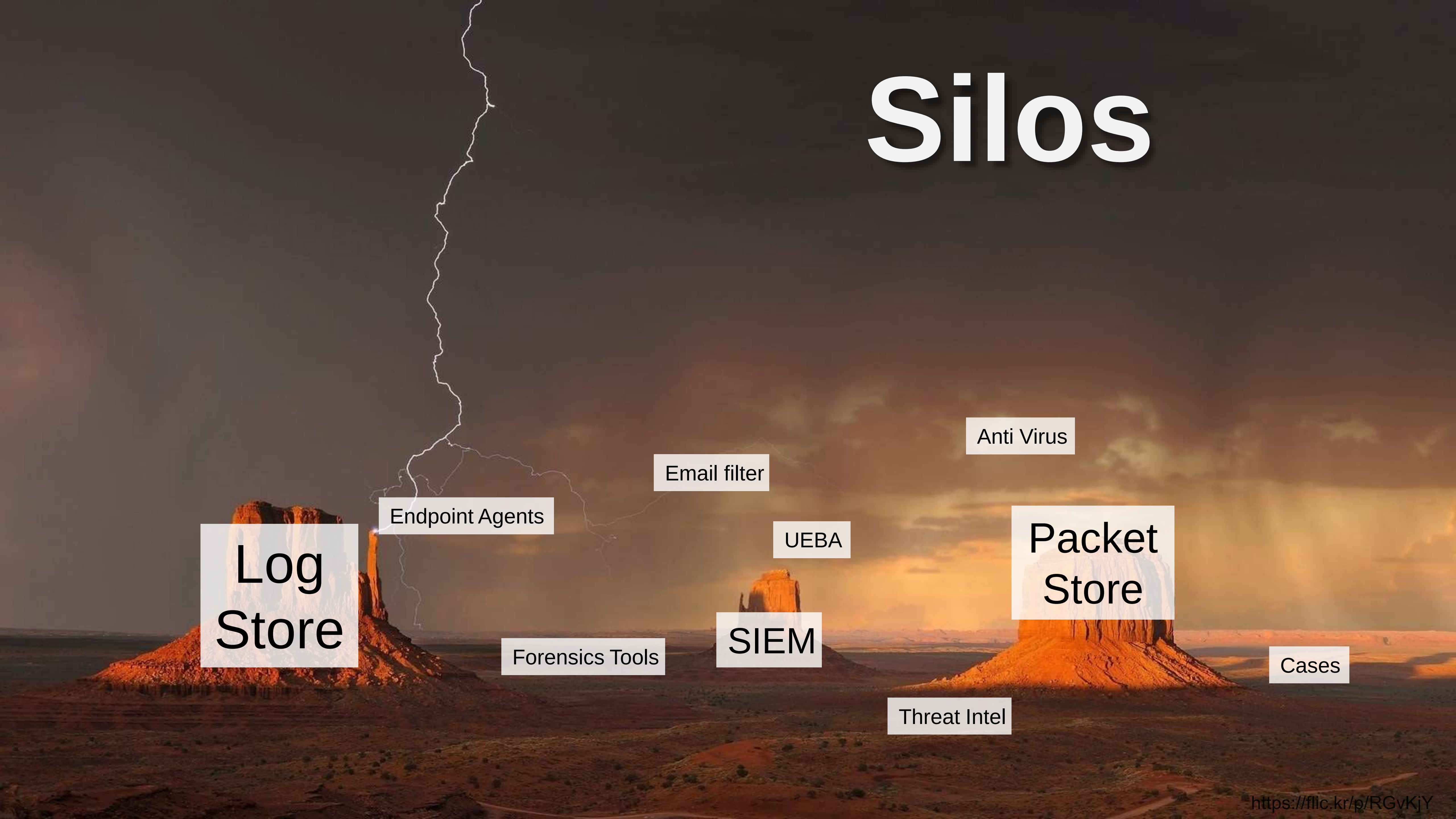
Drowning in Data

Staff shortage

Long tail problem

What we have now

Silos



Anti Virus

Email filter

Endpoint Agents

UEBA

Packet Store

Log Store

SIEM

Forensics Tools

Cases

Threat Intel

https://flic.kr/p/RGvKjY

Rules: Asset or Liability

Shiny new tools

Solutions: machine learning! magic!

Triage Automation

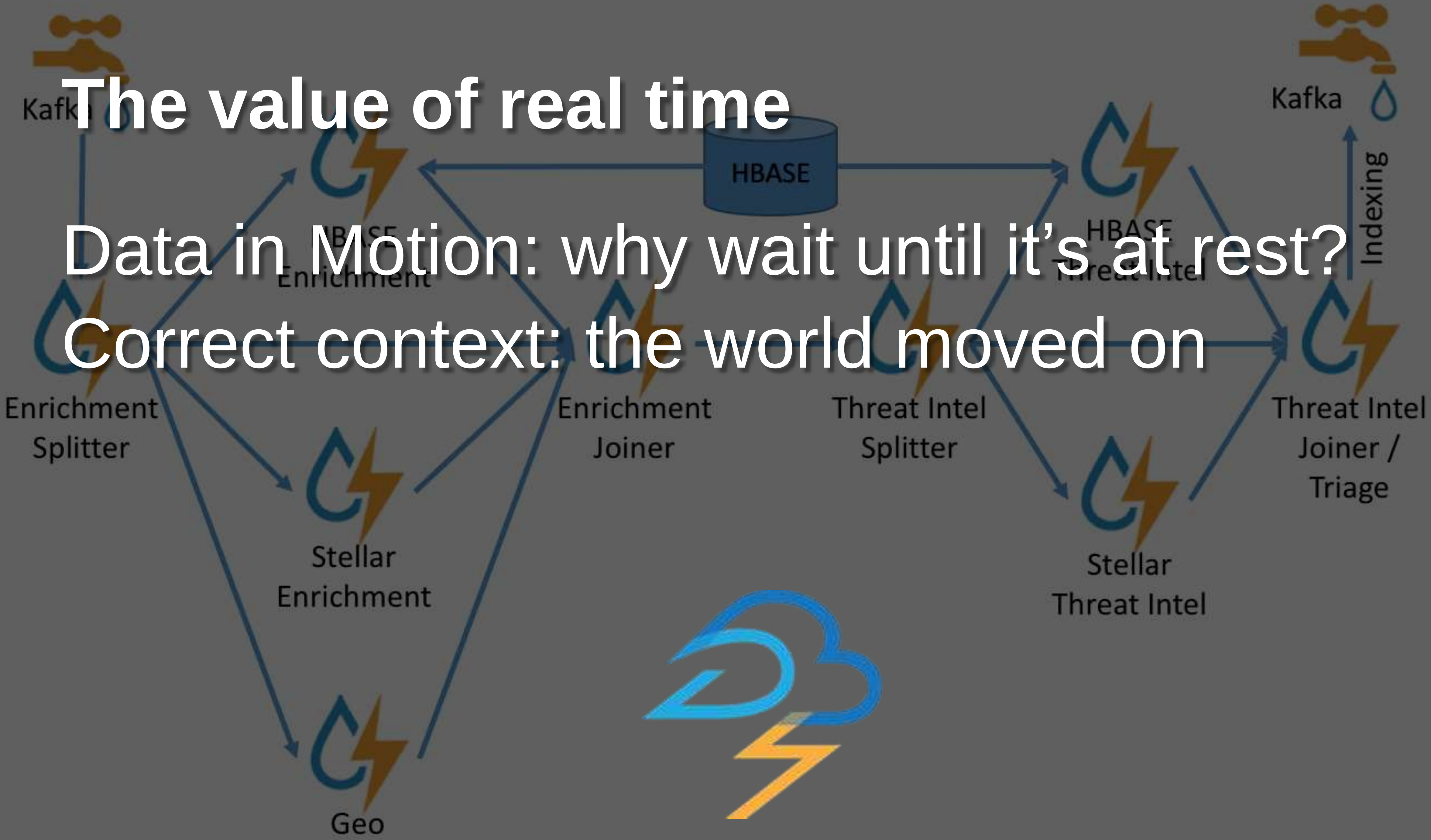
Detecting the unknown unknowns

Explaining yourself

The value of real time

Data in Motion: why wait until it’s at rest?

Correct context: the world moved on



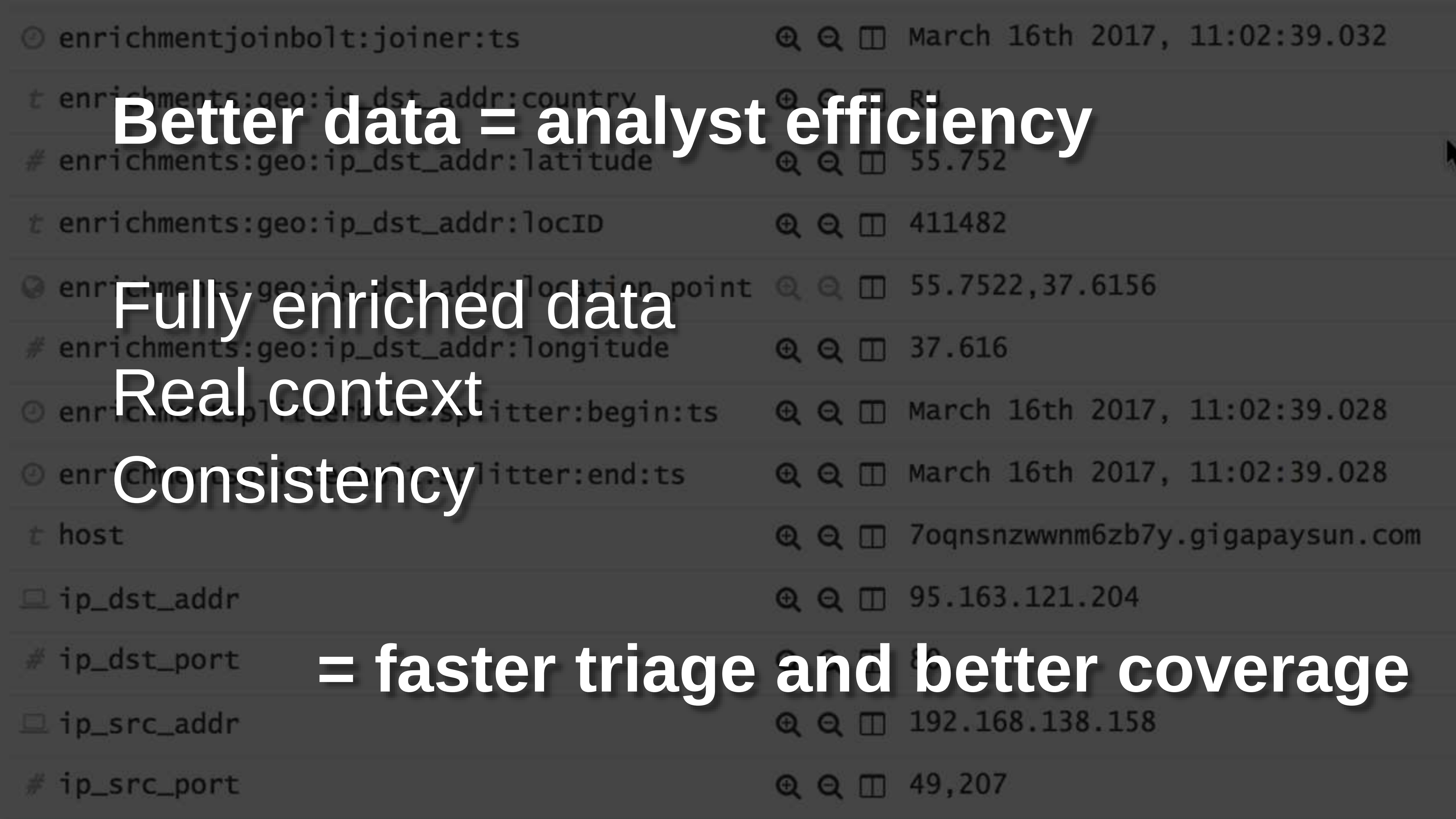
Better data = analyst efficiency

Fully enriched data

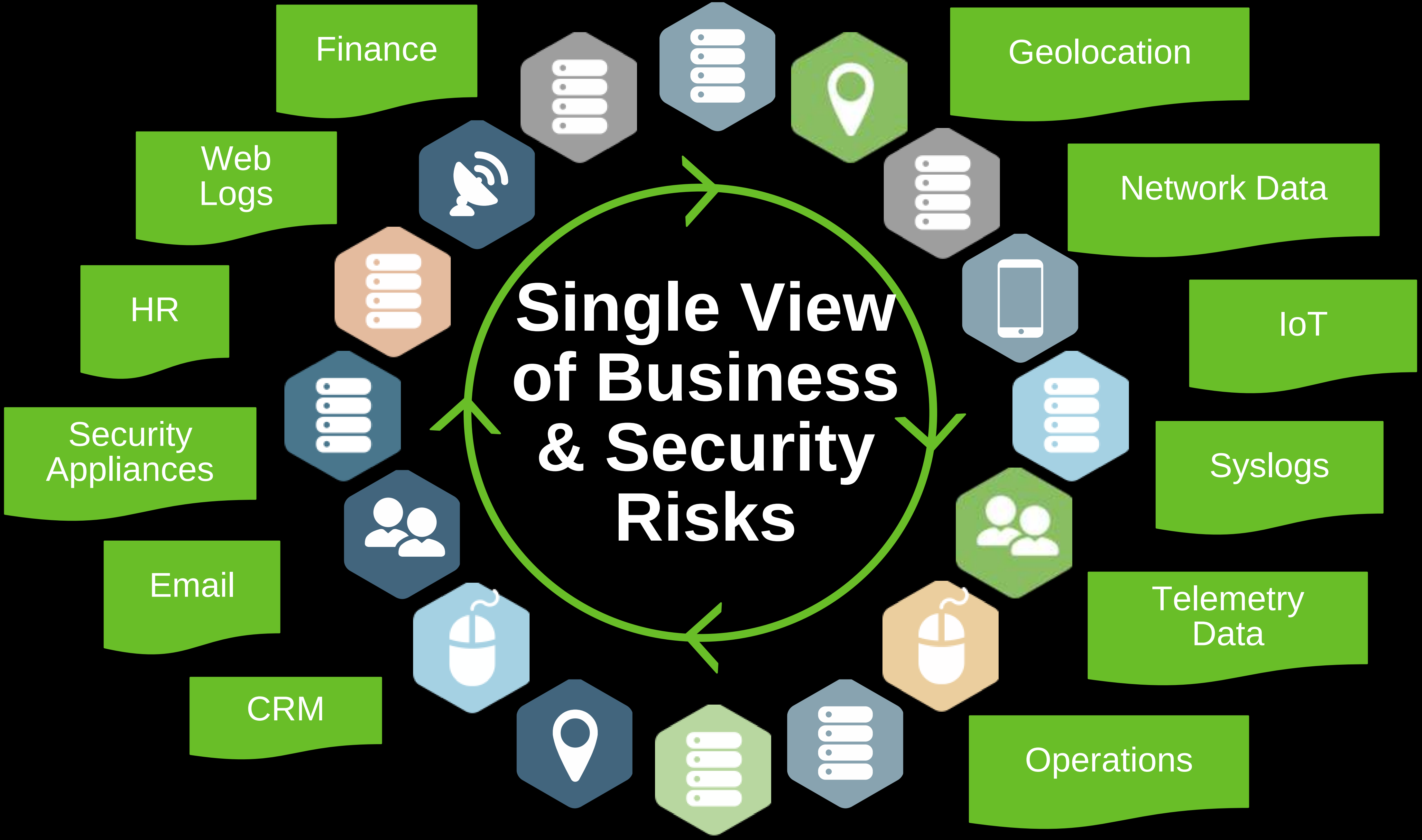
Real context

Consistency

= faster triage and better coverage



enrichmentjoinbolt:joiner:ts	March 16th 2017, 11:02:39.032
enrichments:geo:ip_dst_addr:country	RU
enrichments:geo:ip_dst_addr:latitude	55.752
enrichments:geo:ip_dst_addr:locID	411482
enrichments:geo:ip_dst_addr:location_point	55.7522,37.6156
enrichments:geo:ip_dst_addr:longitude	37.616
enrichmentsplitterbolt:splitter:begin:ts	March 16th 2017, 11:02:39.028
enrichmentsplitterbolt:splitter:end:ts	March 16th 2017, 11:02:39.028
host	7oqnsnzwwnm6zb7y.gigapaysun.com
ip_dst_addr	95.163.121.204
ip_dst_port	80
ip_src_addr	192.168.138.158
ip_src_port	49,207



Single View of Business & Security Risks

Finance

Web Logs

HR

Security Appliances

Email

CRM

Geolocation

Network Data

IoT

Syslogs

Telemetry Data

Operations

Longer term data

- Attacks last months
- So should your queryable data

Executable solutions

- Orchestration
- Machine-time response

How to do it

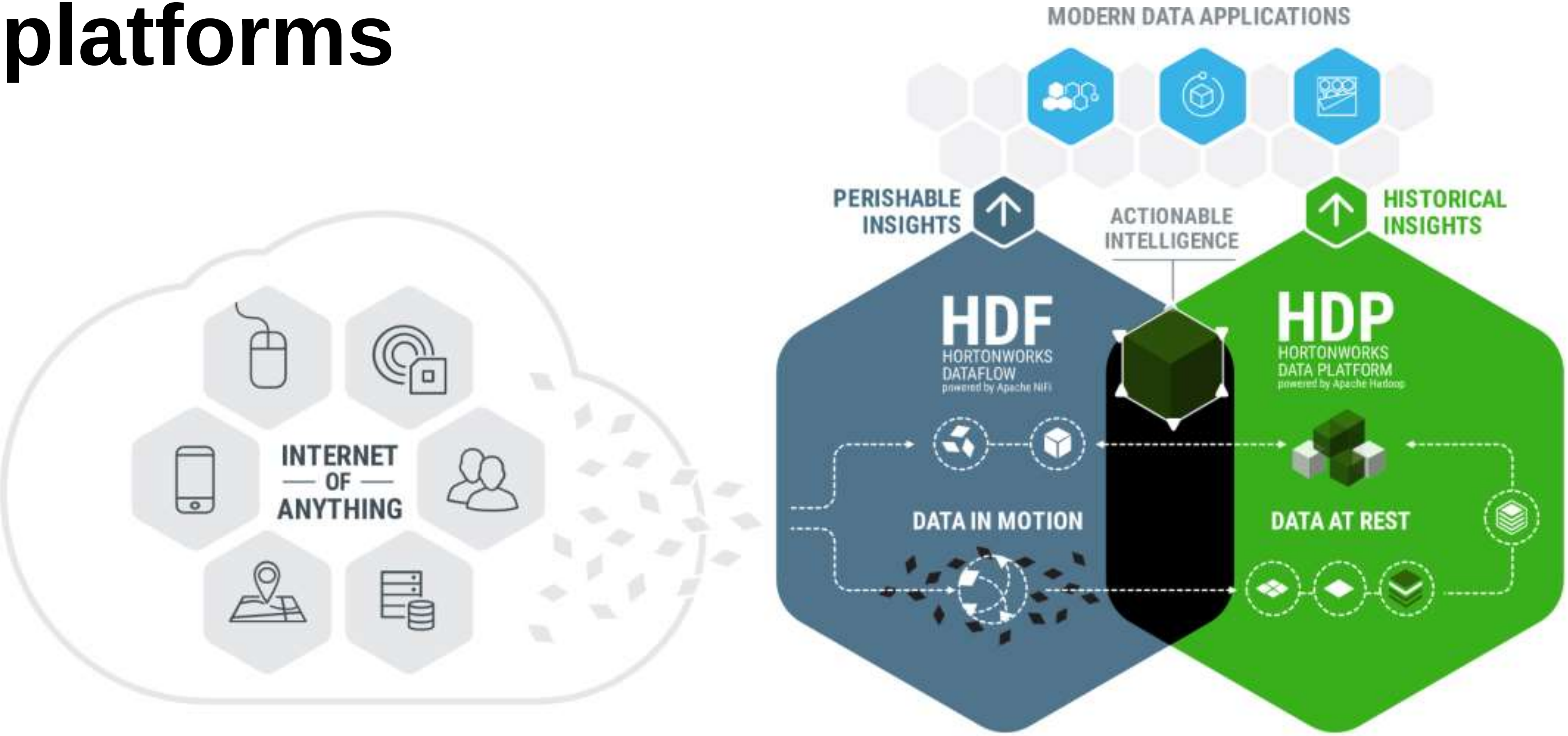
Network Level Taps

Data Sources and Aggregation

Open standards for data models = more productive data scientists + shareable models

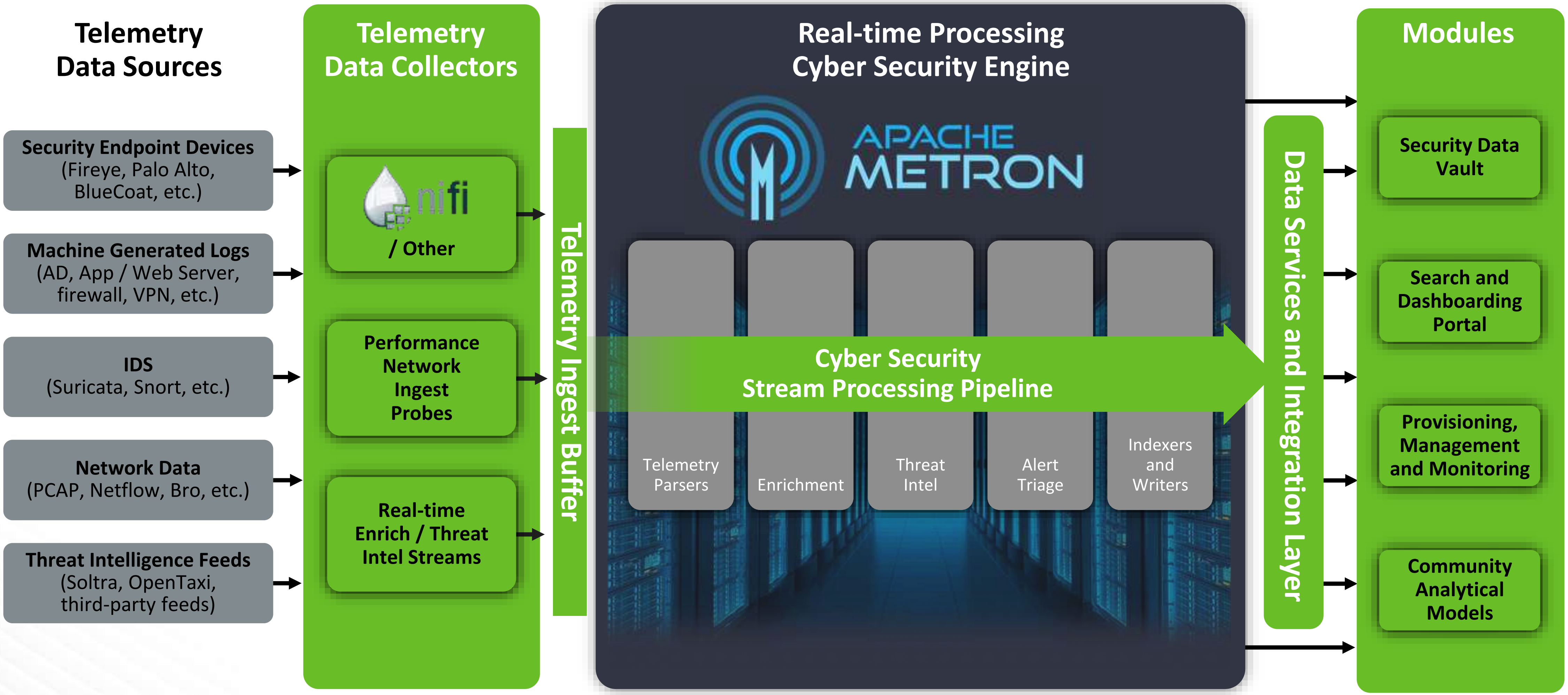
Business level data sources link security to **real business risk**.

Massively scalable platforms



DIY

Apache Metron: a framework for Big Data Driven cyber security



Telemetry Data Sources

- **Security Endpoint Devices** (Fireye, Palo Alto, BlueCoat, etc.)
- **Machine Generated Logs** (AD, App / Web Server, firewall, VPN, etc.)
- **IDS** (Suricata, Snort, etc.)
- **Network Data** (PCAP, Netflow, Bro, etc.)
- **Threat Intelligence Feeds** (Soltra, OpenTaxi, third-party feeds)

Telemetry Data Collectors

- nifi / Other
- Performance Network Ingest Probes
- Real-time Enrich / Threat Intel Streams

Telemetry Ingest Buffer

Real-time Processing Cyber Security Engine

APACHE METRON

Cyber Security Stream Processing Pipeline

- Telemetry Parsers
- Enrichment
- Threat Intel
- Alert Triage
- Indexers and Writers

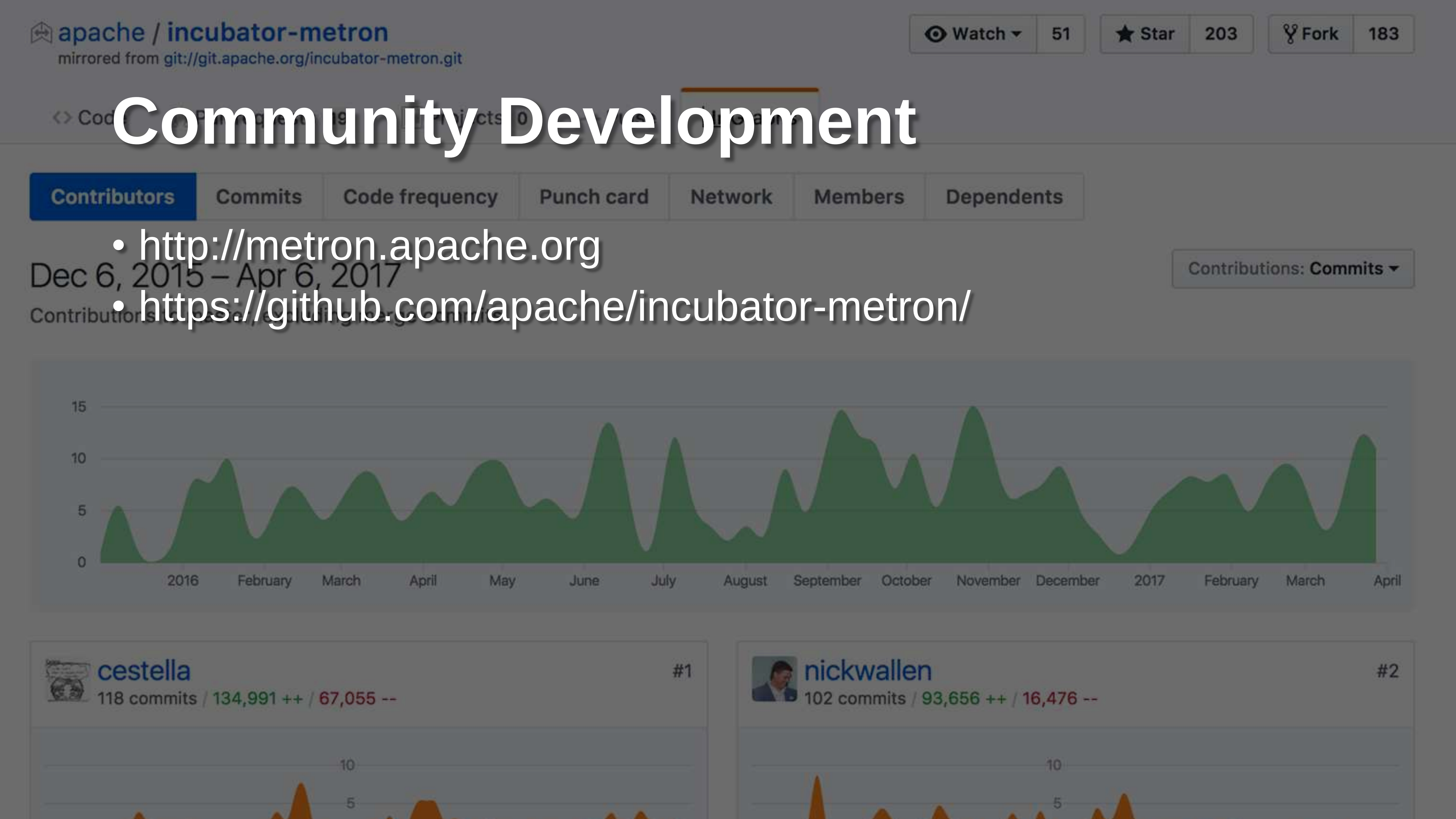
Data Services and Integration Layer

Modules

- Security Data Vault
- Search and Dashboarding Portal
- Provisioning, Management and Monitoring
- Community Analytical Models

Community Development

- http://metron.apache.org
- https://github.com/apache/incubator-metron/



Dec 6, 2015 – Apr 6, 2017

cestella
118 commits / 134,991 ++ / 67,055 --

nickwallen
102 commits / 93,656 ++ / 16,476 --

Thank you!

- Apache Metron: **http://metron.apache.org**

Twitter: **@sireb**