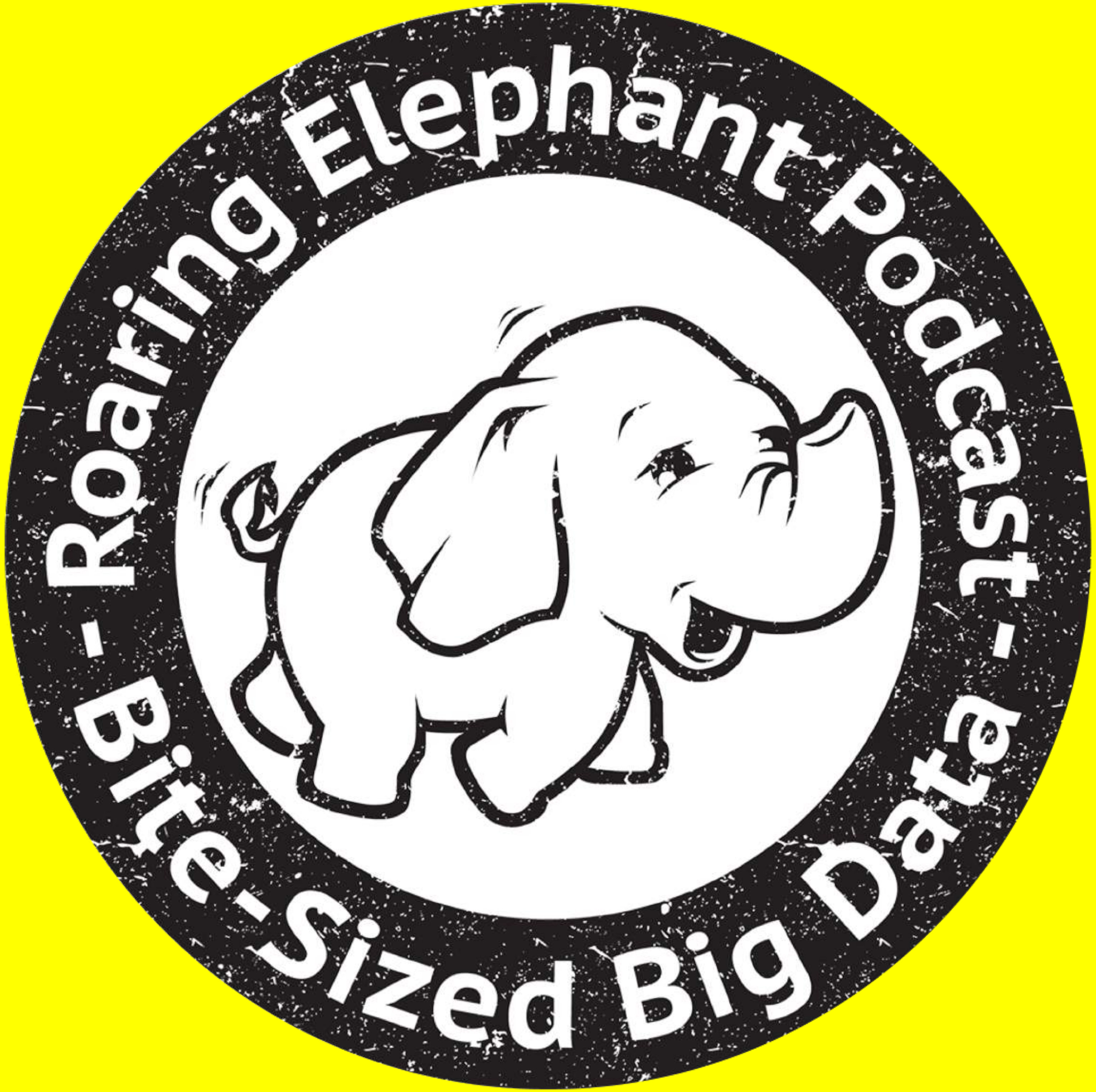
bigdata 27-29 November, Vilnius

Apache Metron in the Real World

Dave Russell - Hortonworks

www.roaringelephant.org

Who am I?


- Roaring Elephant Podcast -
Bite-Sized Big Data

Why Apache Metron?

Months until breach noticed

9

VS

Avg. months log retention

6

Months missing

3

Time until breach actually noticed

Police One/Berkut

28 Months

Yahoo/FSB

35 Months

FB/Cambridge Analytica

48 Months

“Sometime in the next few years we're going to have our first category-one cyber-incident; one that will need a national response.”

Ian Levy
Technical Director
National Cyber Security Centre

Andhra Pradesh Police, India
Aristotle University of Thessaloniki, Greece
Automobile Dacia, Romania
Cambrian College, Canada
Chinese public security bureau
CJ CGV
Dalian Maritime University
Deutsche Bahn
Dharmais Hospital, Indonesia
Faculty Hospital, Nitra, Slovakia
FedEx
Garena Blade and Soul
Guilin University Of Aerospace Technology
Guilin University Of Electronic Technology
Harapan Kita
Hospital[disambiguation needed], IndonesiaHezhou University

Hitachi
Honda
Instituto Nacional de Salud, Colombia
Lakeridge Health
LAKS
LATAM Airlines Group
MegaFon
Ministry of Internal Affairs of the Russian Federation
Ministry of Foreign Affairs (Romania)
National Health Service (England)
NHS Scotland
Nissan Motor Manufacturing UK
O2, Germany
Petrobrás
PetroChina
Portugal Telecom
Pulse FM
Q-Park
Renault
Russian Railways

Sandvik
São Paulo Court of Justice
Saudi Telecom Company
Sberbank
Shandong University
State Governments of India
Government of Gujarat
Government of Kerala
Government of Maharashtra
Government of West Bengal
Suzhou Vehicle Administration
Sun Yat-sen University, China
Telefónica
Telenor Hungary, Hungary
Telkom (South Africa)
Timrå Municipality, Sweden
Universitas Jember, Indonesia
University of Milano-Bicocca, Italy
University of Montreal, Canada
Vivo, Brazil

2018 so far...

EXACTIS

340M Records

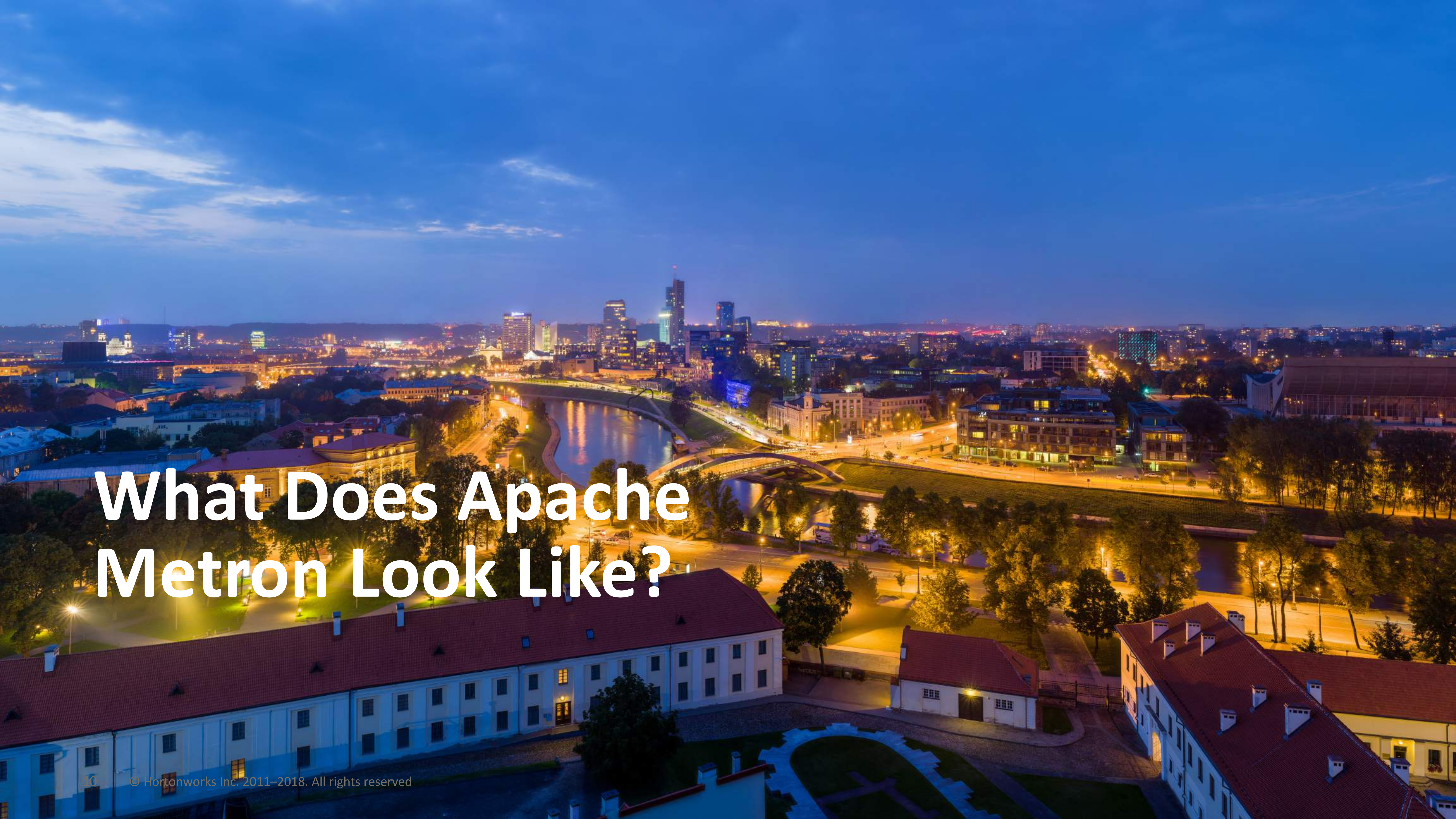
MyHeritage

92M Records

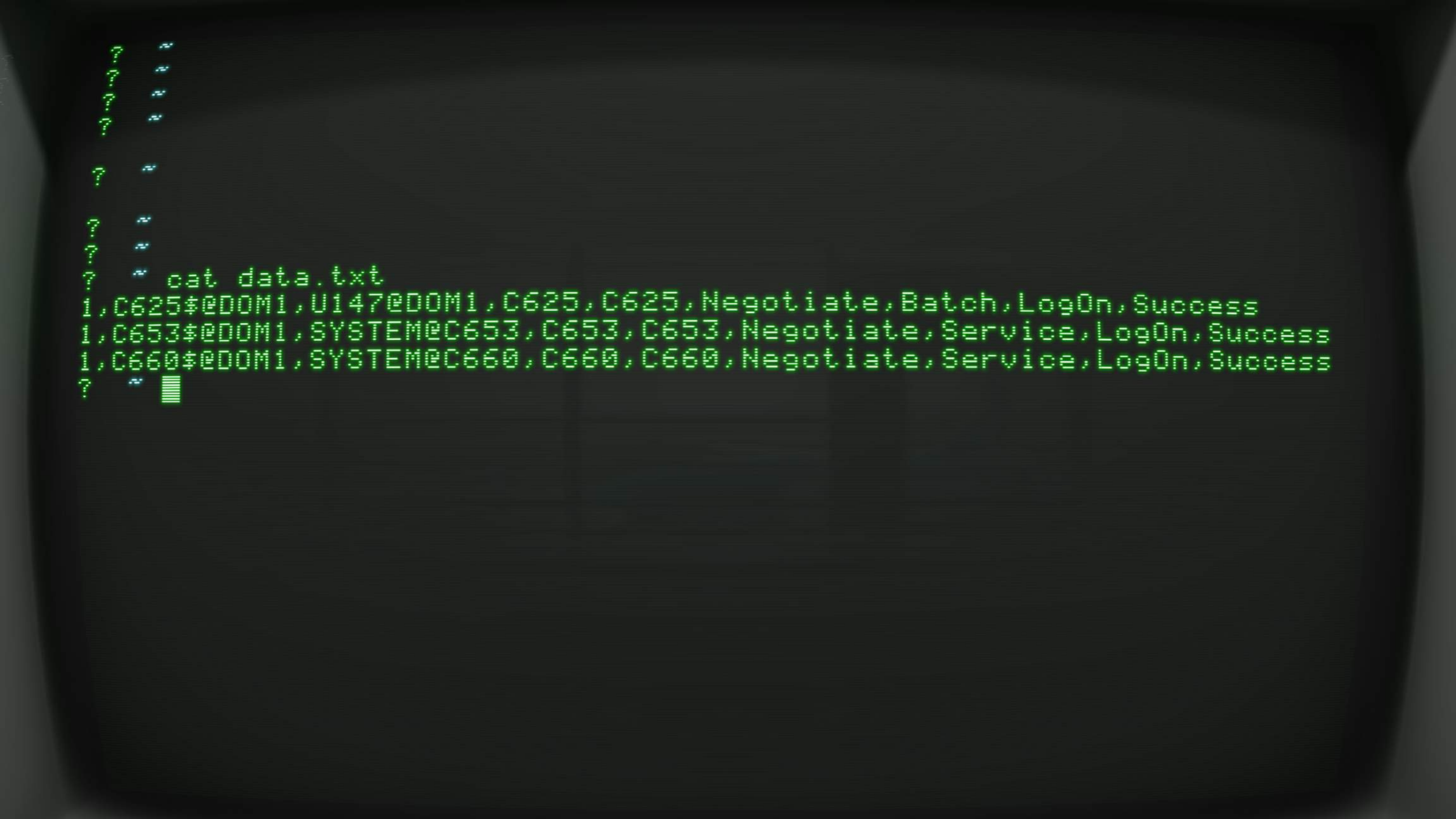
UNDER ARMOUR®

150M Records

And many, many, many more..
https://en.wikipedia.org/wiki/List_of_data_breaches



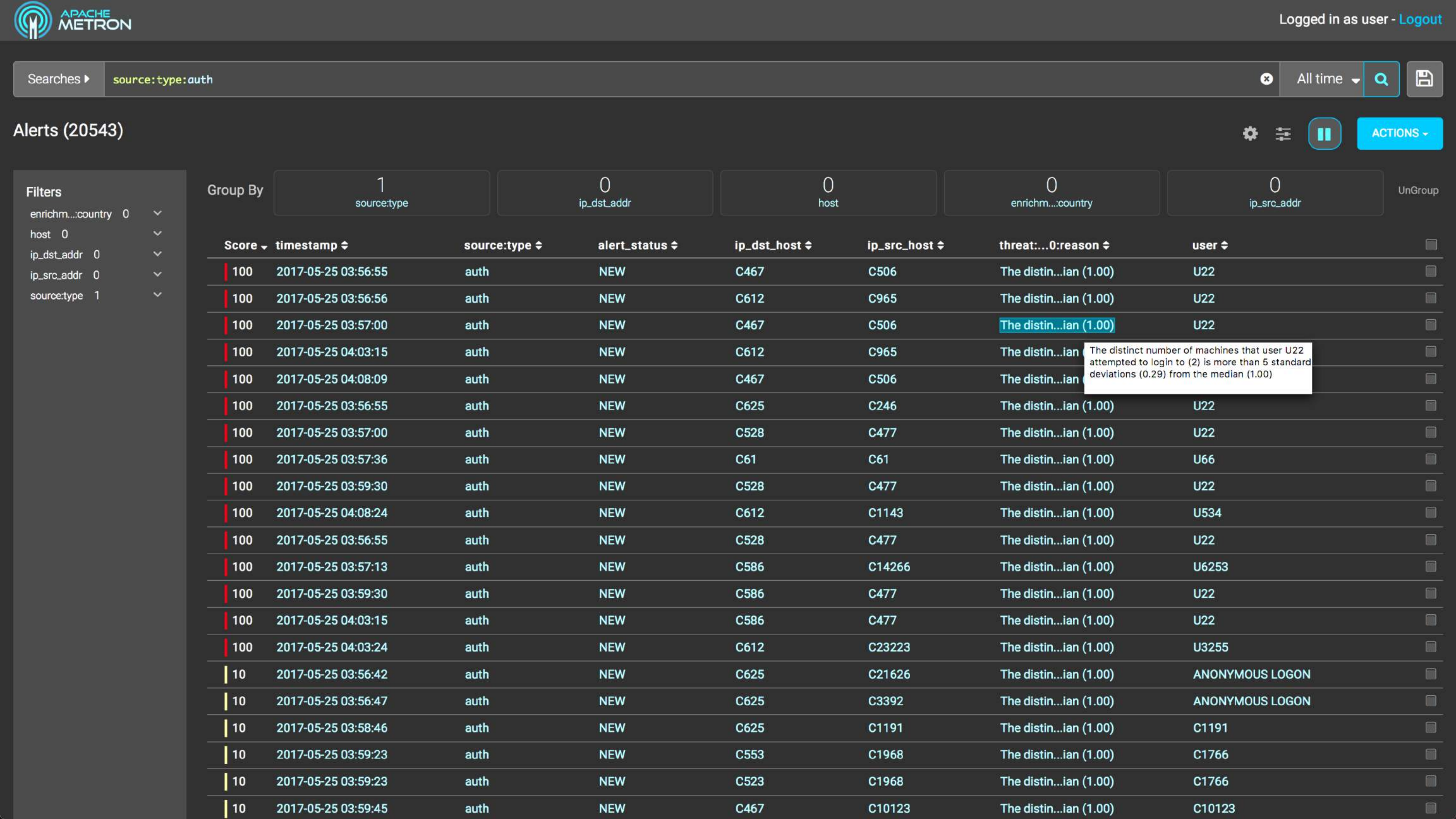
What Does Apache Metron Look Like?



```
?  ~
?  ~
?  ~
?  ~

?  ~

?  ~
?  ~
?  ~  cat data.txt
1,C625$@DOM1,U147@DOM1,C625,C625,Negotiate,Batch,LogOn,Success
1,C653$@DOM1,SYSTEM@C653,C653,C653,Negotiate,Service,LogOn,Success
1,C660$@DOM1,SYSTEM@C660,C660,C660,Negotiate,Service,LogOn,Success
?  ~  █
```



source:type:auth

Alerts (20543)

Filters

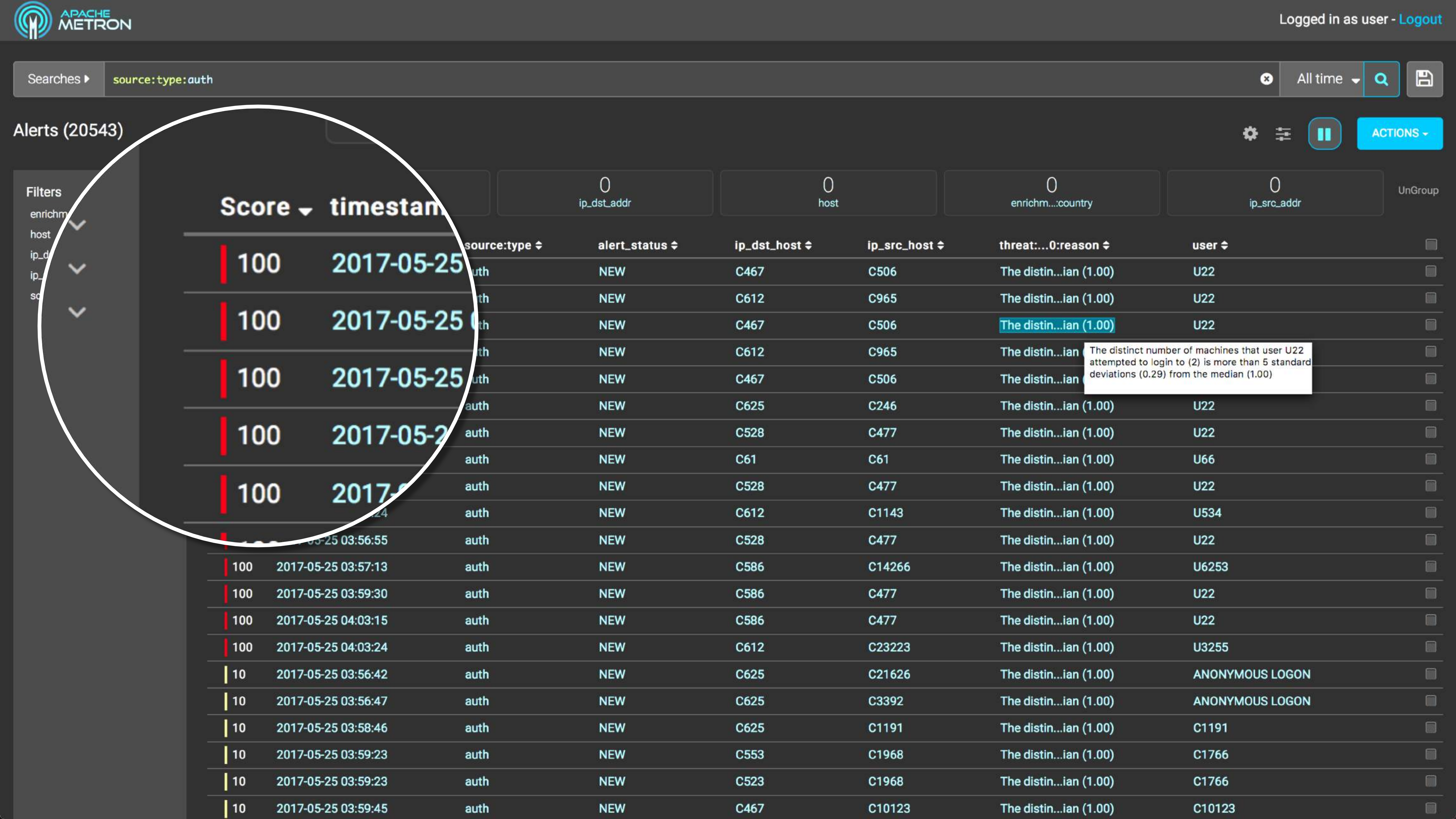
enrichm...:country	0
host	0
ip_dst_addr	0
ip_src_addr	0
source:type	1

Group By

source:type	ip_dst_addr	host	enrichm...:country	ip_src_addr
1	0	0	0	0

Score	timestamp	source:type	alert_status	ip_dst_host	ip_src_host	threat:...0:reason	user
100	2017-05-25 03:56:55	auth	NEW	C467	C506	The distin...ian (1.00)	U22
100	2017-05-25 03:56:56	auth	NEW	C612	C965	The distin...ian (1.00)	U22
100	2017-05-25 03:57:00	auth	NEW	C467	C506	The distin...ian (1.00)	U22
100	2017-05-25 04:03:15	auth	NEW	C612	C965	The distin...ian (	
100	2017-05-25 04:08:09	auth	NEW	C467	C506	The distin...ian (	
100	2017-05-25 03:56:55	auth	NEW	C625	C246	The distin...ian (1.00)	U22
100	2017-05-25 03:57:00	auth	NEW	C528	C477	The distin...ian (1.00)	U22
100	2017-05-25 03:57:36	auth	NEW	C61	C61	The distin...ian (1.00)	U66
100	2017-05-25 03:59:30	auth	NEW	C528	C477	The distin...ian (1.00)	U22
100	2017-05-25 04:08:24	auth	NEW	C612	C1143	The distin...ian (1.00)	U534
100	2017-05-25 03:56:55	auth	NEW	C528	C477	The distin...ian (1.00)	U22
100	2017-05-25 03:57:13	auth	NEW	C586	C14266	The distin...ian (1.00)	U6253
100	2017-05-25 03:59:30	auth	NEW	C586	C477	The distin...ian (1.00)	U22
100	2017-05-25 04:03:15	auth	NEW	C586	C477	The distin...ian (1.00)	U22
100	2017-05-25 04:03:24	auth	NEW	C612	C23223	The distin...ian (1.00)	U3255
10	2017-05-25 03:56:42	auth	NEW	C625	C21626	The distin...ian (1.00)	ANONYMOUS LOGON
10	2017-05-25 03:56:47	auth	NEW	C625	C3392	The distin...ian (1.00)	ANONYMOUS LOGON
10	2017-05-25 03:58:46	auth	NEW	C625	C1191	The distin...ian (1.00)	C1191
10	2017-05-25 03:59:23	auth	NEW	C553	C1968	The distin...ian (1.00)	C1766
10	2017-05-25 03:59:23	auth	NEW	C523	C1968	The distin...ian (1.00)	C1766
10	2017-05-25 03:59:45	auth	NEW	C467	C10123	The distin...ian (1.00)	C10123

The distinct number of machines that user U22 attempted to login to (2) is more than 5 standard deviations (0.29) from the median (1.00)



source:type:auth

Alerts (20543)

0	0	0	0
ip_dst_addr	host	enrichm...:country	ip_src_addr

Score ▾ timestam

100 2017-05-25

100 2017-05-25

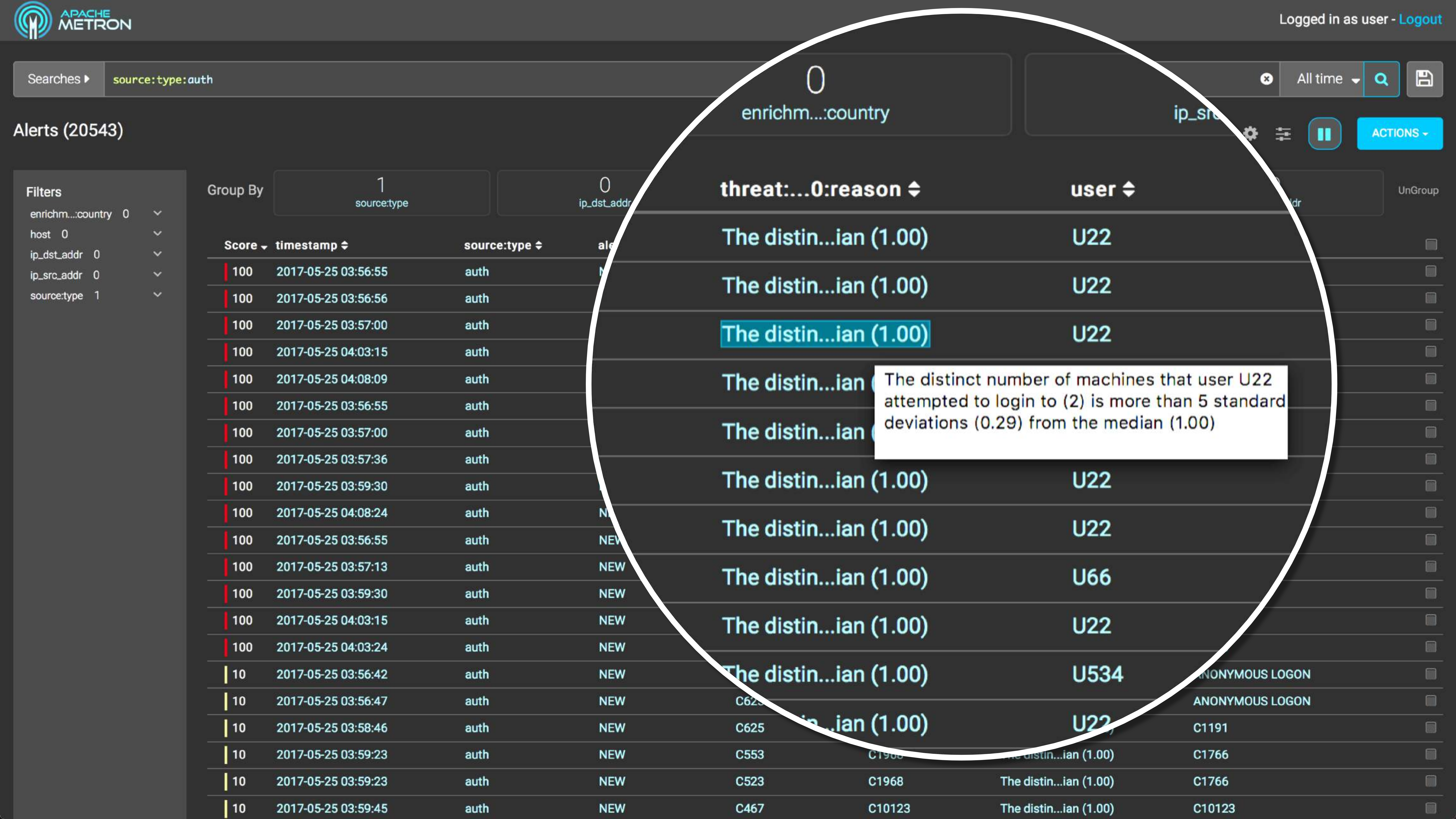
100 2017-05-25

100 2017-05-2

100 2017-

Score	timestamp	source:type ⇕	alert_status ⇕	ip_dst_host ⇕	ip_src_host ⇕	threat:...0:reason ⇕	user ⇕
		uth	NEW	C467	C506	The distin...ian (1.00)	U22
		th	NEW	C612	C965	The distin...ian (1.00)	U22
		th	NEW	C467	C506	The distin...ian (1.00)	U22
		th	NEW	C612	C965	The distin...ian	
		uth	NEW	C467	C506	The distin...ian	
		auth	NEW	C625	C246	The distin...ian (1.00)	U22
		auth	NEW	C528	C477	The distin...ian (1.00)	U22
		auth	NEW	C61	C61	The distin...ian (1.00)	U66
		auth	NEW	C528	C477	The distin...ian (1.00)	U22
	24	auth	NEW	C612	C1143	The distin...ian (1.00)	U534
	05-25 03:56:55	auth	NEW	C528	C477	The distin...ian (1.00)	U22
100	2017-05-25 03:57:13	auth	NEW	C586	C14266	The distin...ian (1.00)	U6253
100	2017-05-25 03:59:30	auth	NEW	C586	C477	The distin...ian (1.00)	U22
100	2017-05-25 04:03:15	auth	NEW	C586	C477	The distin...ian (1.00)	U22
100	2017-05-25 04:03:24	auth	NEW	C612	C23223	The distin...ian (1.00)	U3255
10	2017-05-25 03:56:42	auth	NEW	C625	C21626	The distin...ian (1.00)	ANONYMOUS LOGON
10	2017-05-25 03:56:47	auth	NEW	C625	C3392	The distin...ian (1.00)	ANONYMOUS LOGON
10	2017-05-25 03:58:46	auth	NEW	C625	C1191	The distin...ian (1.00)	C1191
10	2017-05-25 03:59:23	auth	NEW	C553	C1968	The distin...ian (1.00)	C1766
10	2017-05-25 03:59:23	auth	NEW	C523	C1968	The distin...ian (1.00)	C1766
10	2017-05-25 03:59:45	auth	NEW	C467	C10123	The distin...ian (1.00)	C10123

The distinct number of machines that user U22 attempted to login to (2) is more than 5 standard deviations (0.29) from the median (1.00)



source:type:auth

Alerts (20543)

Filters

enrichm...:country	0
host	0
ip_dst_addr	0
ip_src_addr	0
source:type	1

Group By

1	0	0
source:type	ip_dst_addr	enrichm...:country

Score	timestamp	source:type	ale
100	2017-05-25 03:56:55	auth	N
100	2017-05-25 03:56:56	auth	
100	2017-05-25 03:57:00	auth	
100	2017-05-25 04:03:15	auth	
100	2017-05-25 04:08:09	auth	
100	2017-05-25 03:56:55	auth	
100	2017-05-25 03:57:00	auth	
100	2017-05-25 03:57:36	auth	
100	2017-05-25 03:59:30	auth	
100	2017-05-25 04:08:24	auth	N
100	2017-05-25 03:56:55	auth	NEW
100	2017-05-25 03:57:13	auth	NEW
100	2017-05-25 03:59:30	auth	NEW
100	2017-05-25 04:03:15	auth	NEW
100	2017-05-25 04:03:24	auth	NEW
10	2017-05-25 03:56:42	auth	NEW
10	2017-05-25 03:56:47	auth	NEW
10	2017-05-25 03:58:46	auth	NEW
10	2017-05-25 03:59:23	auth	NEW
10	2017-05-25 03:59:23	auth	NEW
10	2017-05-25 03:59:45	auth	NEW

			NONYMOUS LOGON
C62			ANONYMOUS LOGON
C625			C1191
C553	C1968	The distin...ian (1.00)	C1766
C523	C1968	The distin...ian (1.00)	C1766
C467	C10123	The distin...ian (1.00)	C10123

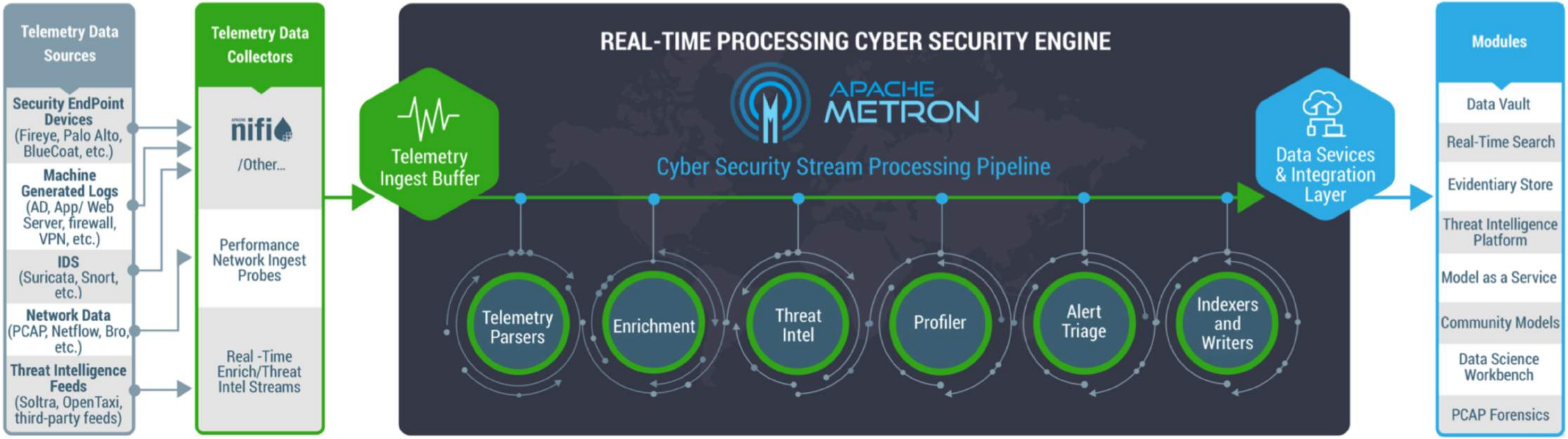
threat:...0:reason	user
The distin...ian (1.00)	U22
The distin...ian (1.00)	U22
The distin...ian (1.00)	U22
The distin...ian	
The distin...ian	
The distin...ian (1.00)	U22
The distin...ian (1.00)	U22
The distin...ian (1.00)	U66
The distin...ian (1.00)	U22
The distin...ian (1.00)	U534
...ian (1.00)	U22

The distinct number of machines that user U22 attempted to login to (2) is more than 5 standard deviations (0.29) from the median (1.00)

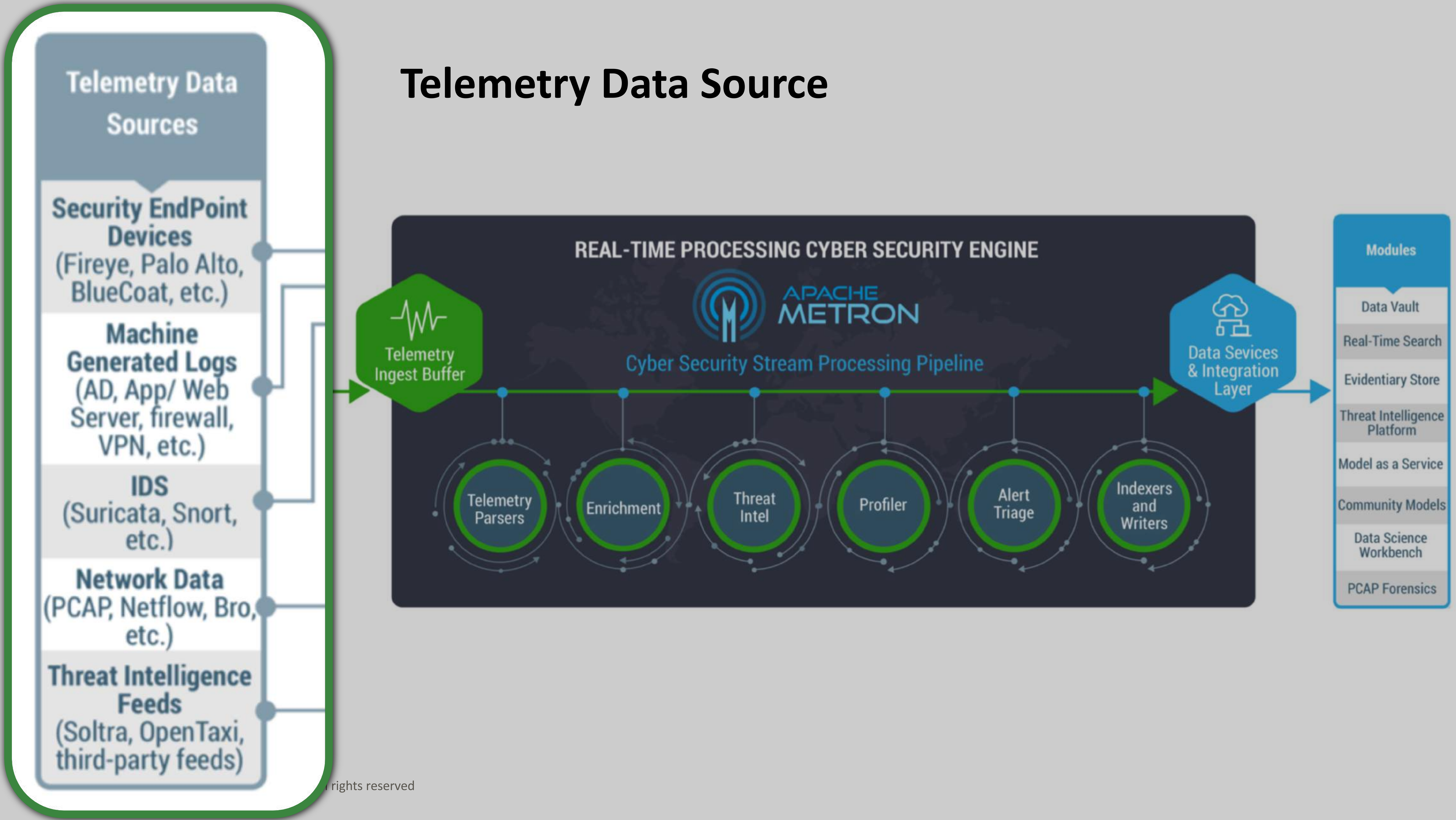
What is Apache Metron?

Built on top on proven open source big data technology

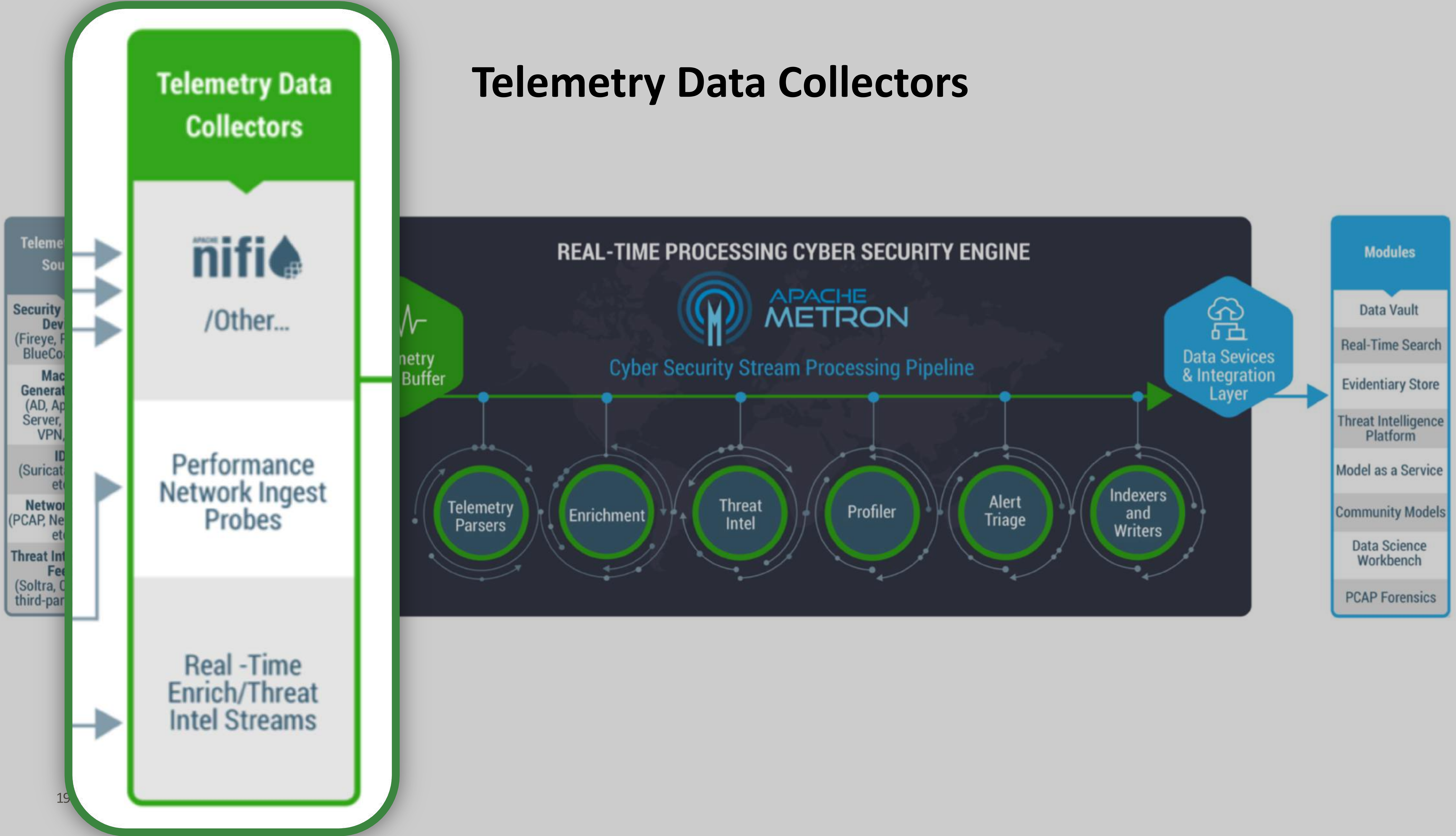
An architecture for real-time cybersecurity analytics


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
REAL-TIME PROCESSING CYBER SECURITY ENGINE
APACHE METRON
Cyber Security Stream Processing Pipeline
Telemetry Ingest Buffer
Telemetry Parsers
Enrichment
Threat Intel
Profiler
Alert Triage
Indexers and Writers
Data Sevices & Integration Layer
Modules
Data Vault
Real-Time Search
Evidentiary Store
Threat Intelligence Platform
Model as a Service
Community Models
Data Science Workbench
PCAP Forensics

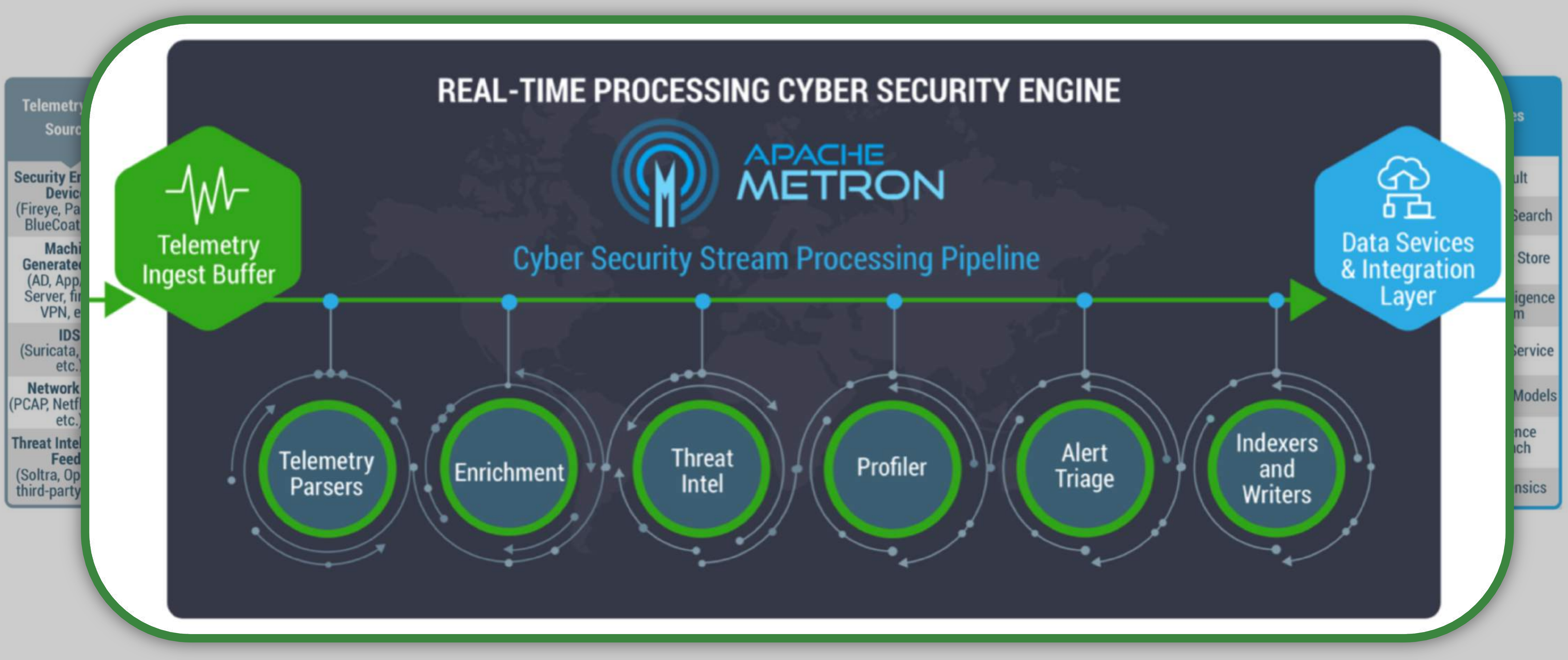
Telemetry Data Source


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)
REAL-TIME PROCESSING CYBER SECURITY ENGINE
APACHE METRON
Cyber Security Stream Processing Pipeline
Telemetry Ingest Buffer
Telemetry Parsers
Enrichment
Threat Intel
Profiler
Alert Triage
Indexers and Writers
Data Sevices & Integration Layer
Modules
Data Vault
Real-Time Search
Evidentiary Store
Threat Intelligence Platform
Model as a Service
Community Models
Data Science Workbench
PCAP Forensics

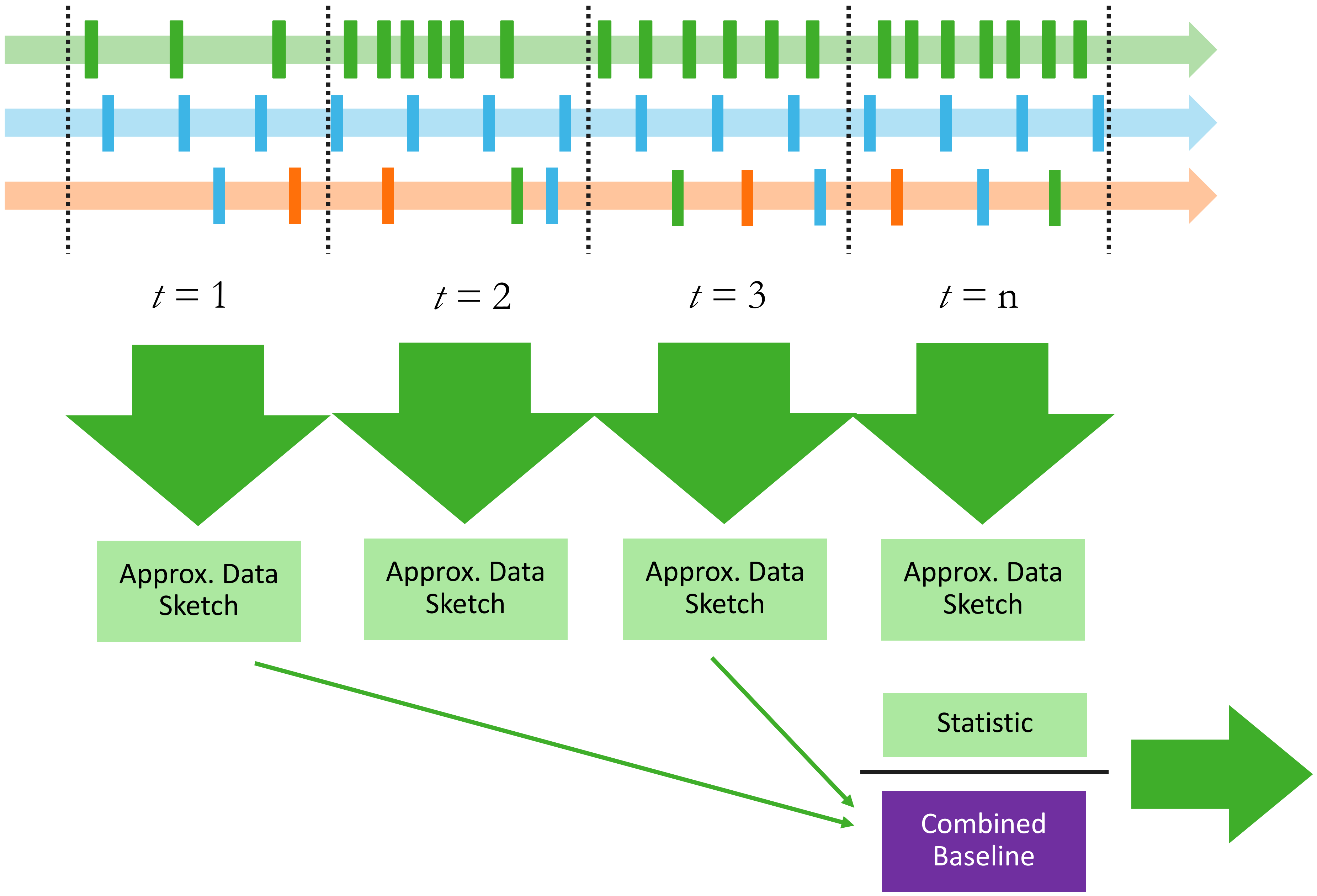
Telemetry Data Collectors



Cyber Security Stream Processing Pipeline


REAL-TIME PROCESSING CYBER SECURITY ENGINE
APACHE METRON
Cyber Security Stream Processing Pipeline
Telemetry Ingest Buffer
Telemetry Parsers
Enrichment
Threat Intel
Profiler
Alert Triage
Indexers and Writers
Data Sevices & Integration Layer
Telemetry Sourc
Security En Devic (Fireye, Pa BlueCoat
Machi Generated (AD, App Server, fi VPN, e
IDS (Suricata, etc.
Network (PCAP, Netfl etc.
Threat Intel Feed (Soltra, Op third-party
ult
Search
Store
igence m
Service
Models
nce ch
nsics

Profiling by time



$t = 1$ $t = 2$ $t = 3$ $t = \mathrm{n}$

Approx. Data Sketch

Approx. Data Sketch

Approx. Data Sketch

Approx. Data Sketch

Statistic

Combined Baseline

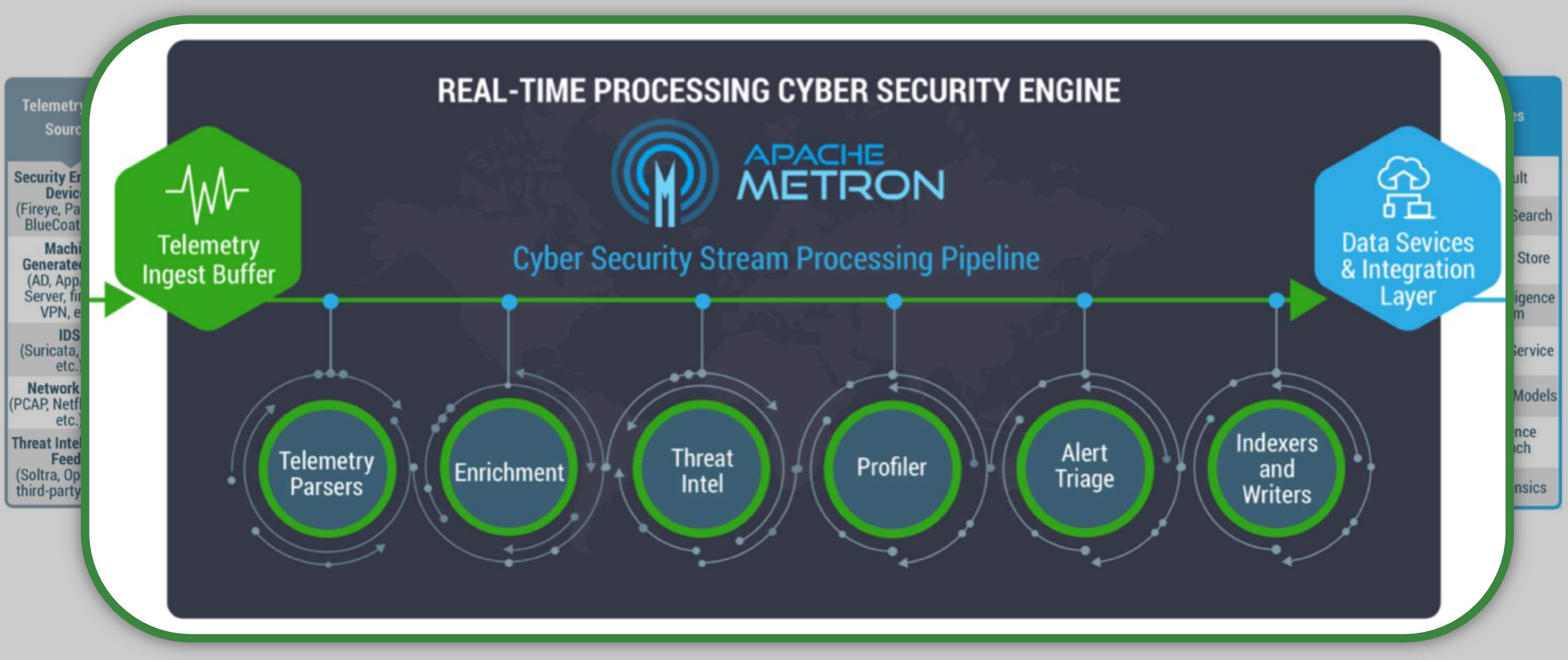
Wide range of algorithms including:

- HyperLogLogPlus
- Bloom filters
- T-digests
- Statistical Baselining
- Hashing functions
- Outlier detection
- GeoHashing over time
- Locality Sensitive Hashing

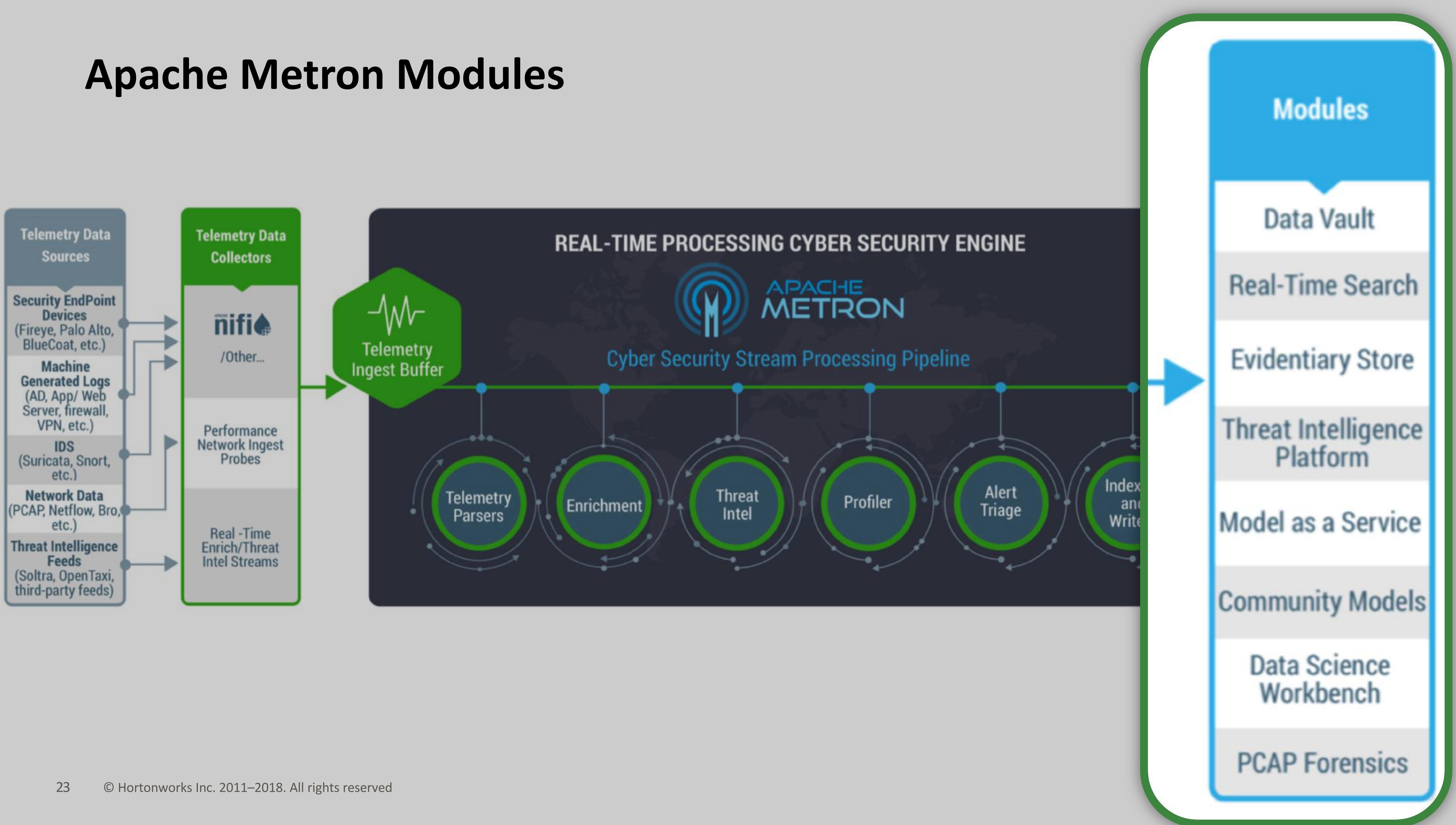
STORM APACHE HBASE

kafka

Cyber Security Stream Processing Pipeline



Apache Metron Modules

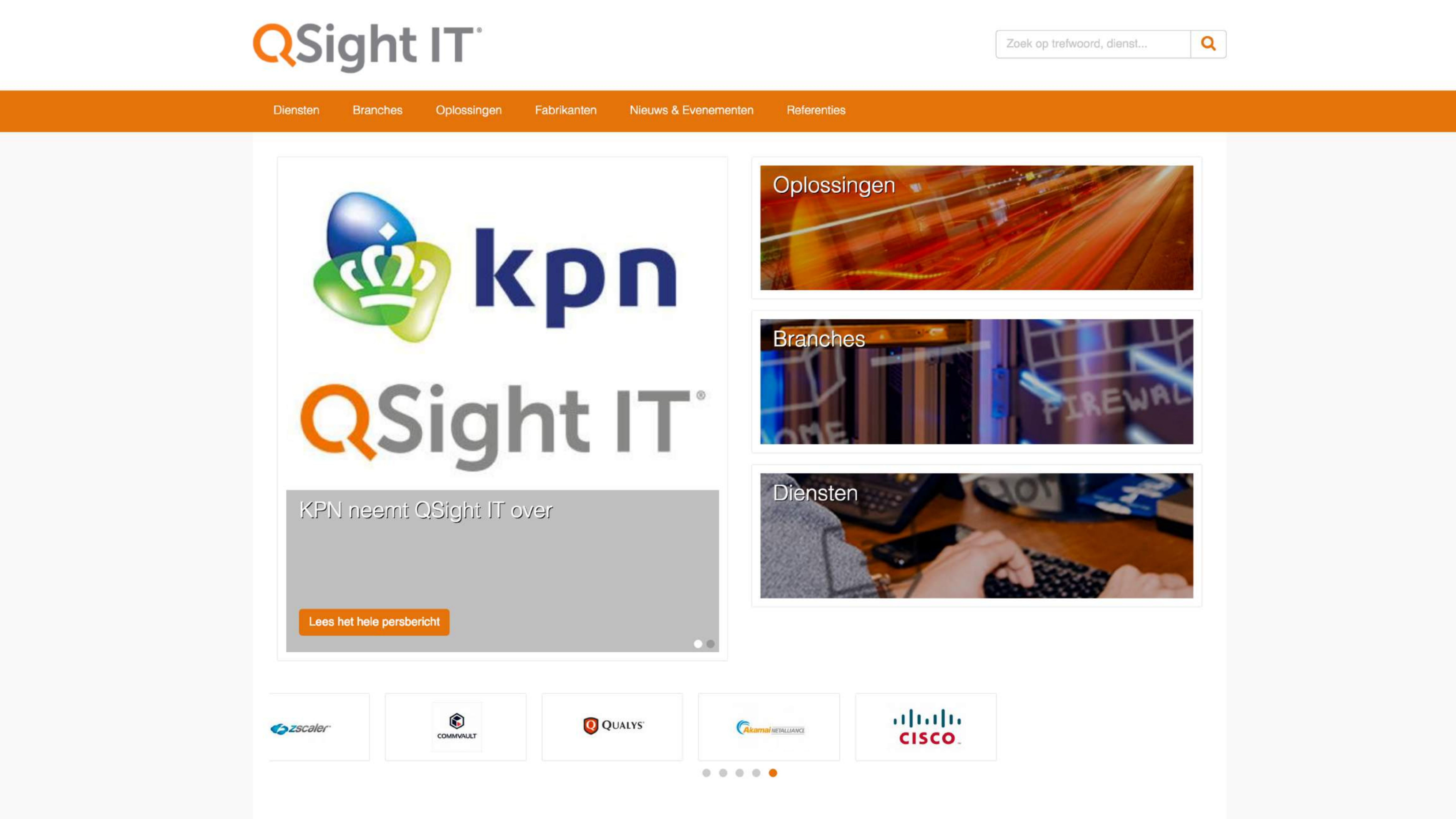

Telemetry Data Sources
Security EndPoint Devices (Fireye, Palo Alto, BlueCoat, etc.)
Machine Generated Logs (AD, App/ Web Server, firewall, VPN, etc.)
IDS (Suricata, Snort, etc.)
Network Data (PCAP, Netflow, Bro, etc.)
Threat Intelligence Feeds (Soltra, OpenTaxi, third-party feeds)
Telemetry Data Collectors
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
Profiler
Alert Triage
Index an Writ
Modules
Data Vault
Real-Time Search
Evidentiary Store
Threat Intelligence Platform
Model as a Service
Community Models
Data Science Workbench
PCAP Forensics

Who is Using Apache Metron (Part 1)

TELSTRA

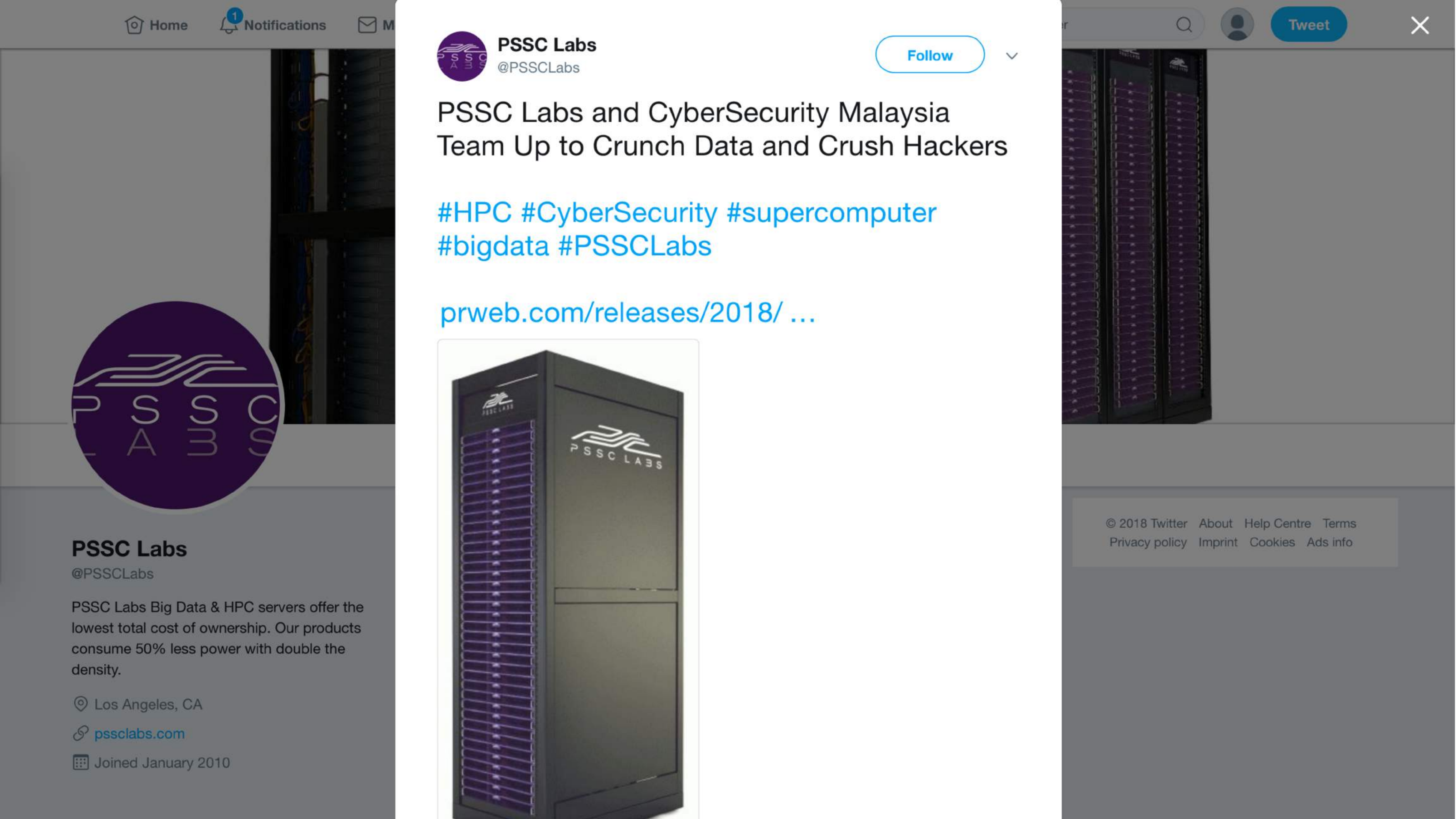
ge_ID: edge:Package-941d9f43-5ed1-4bd7-90bd-de3054675f3a
richmentsplitterbolt:spl end:ts: 146
imestamp: June 27th 2016 57:32.563 ID: fsisac:observabl
27th 2016, 17:57:14. isac:observable-f0d264D9-a6f
2016 06
Capital One
theCUBE
2016 HADOOP SUMMIT

Live from Hadoop Summit 2016 #HS16SJ
San Jose Convention Center - San Jose, CA



KPN neemt QSight IT over

The Wider Apache Metron Ecosystem



PSSC Labs
@PSSCLabs

PSSC Labs and CyberSecurity Malaysia Team Up to Crunch Data and Crush Hackers

#HPC #CyberSecurity #supercomputer #bigdata #PSSCLabs

prweb.com/releases/2018/ …

PSSC Labs
@PSSCLabs

PSSC Labs Big Data & HPC servers offer the lowest total cost of ownership. Our products consume 50% less power with double the density.

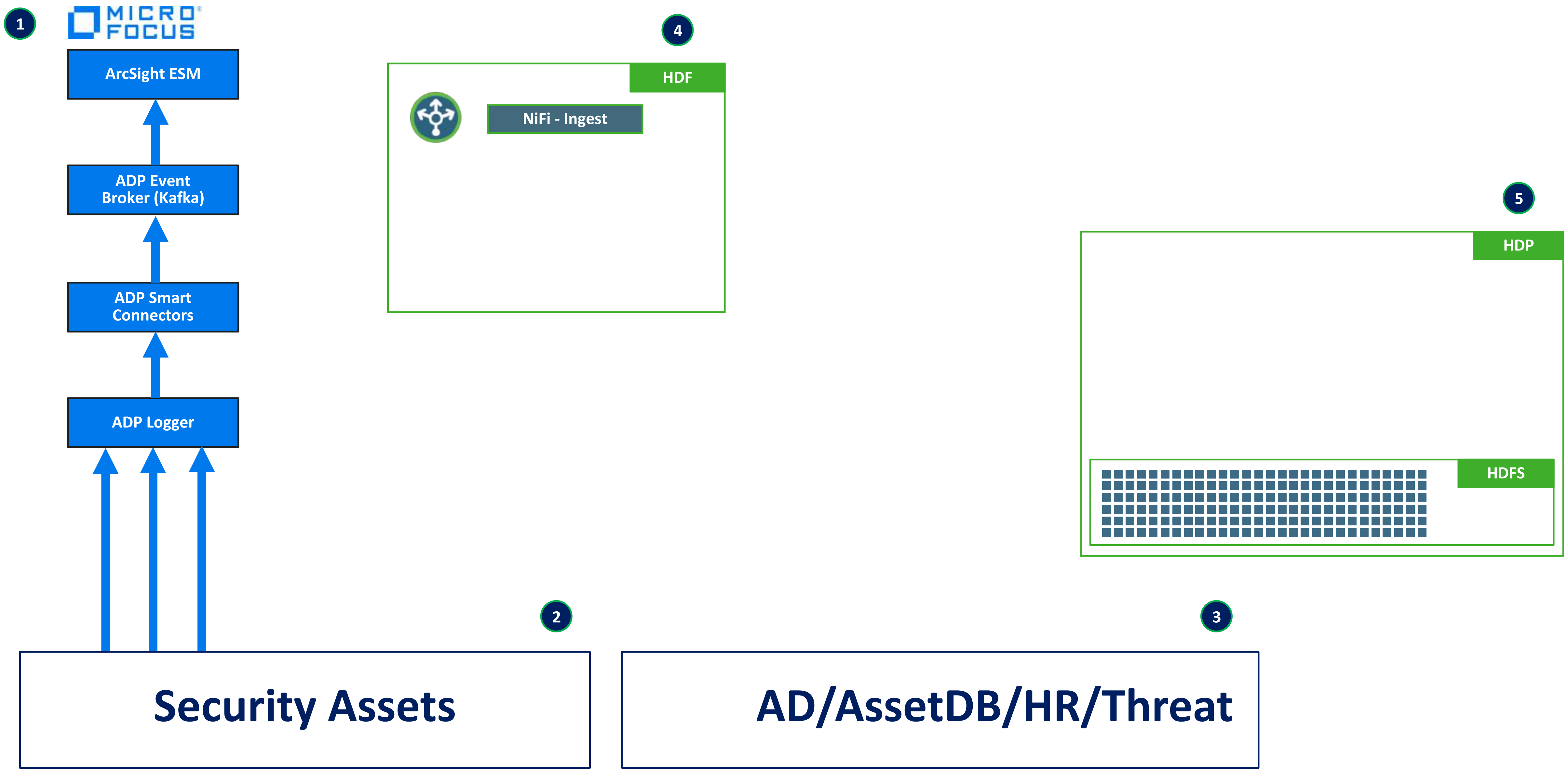
Los Angeles, CA
pssclabs.com
Joined January 2010

© 2018 Twitter About Help Centre Terms
Privacy policy Imprint Cookies Ads info

Who is Using Apache Metron (Part 2)

Deploying Apache Metron

Phase 0 – Current State



1

ArcSight ESM

ADP Event Broker (Kafka)

ADP Smart Connectors

ADP Logger

4

HDF

NiFi - Ingest

5

HDP

HDFS

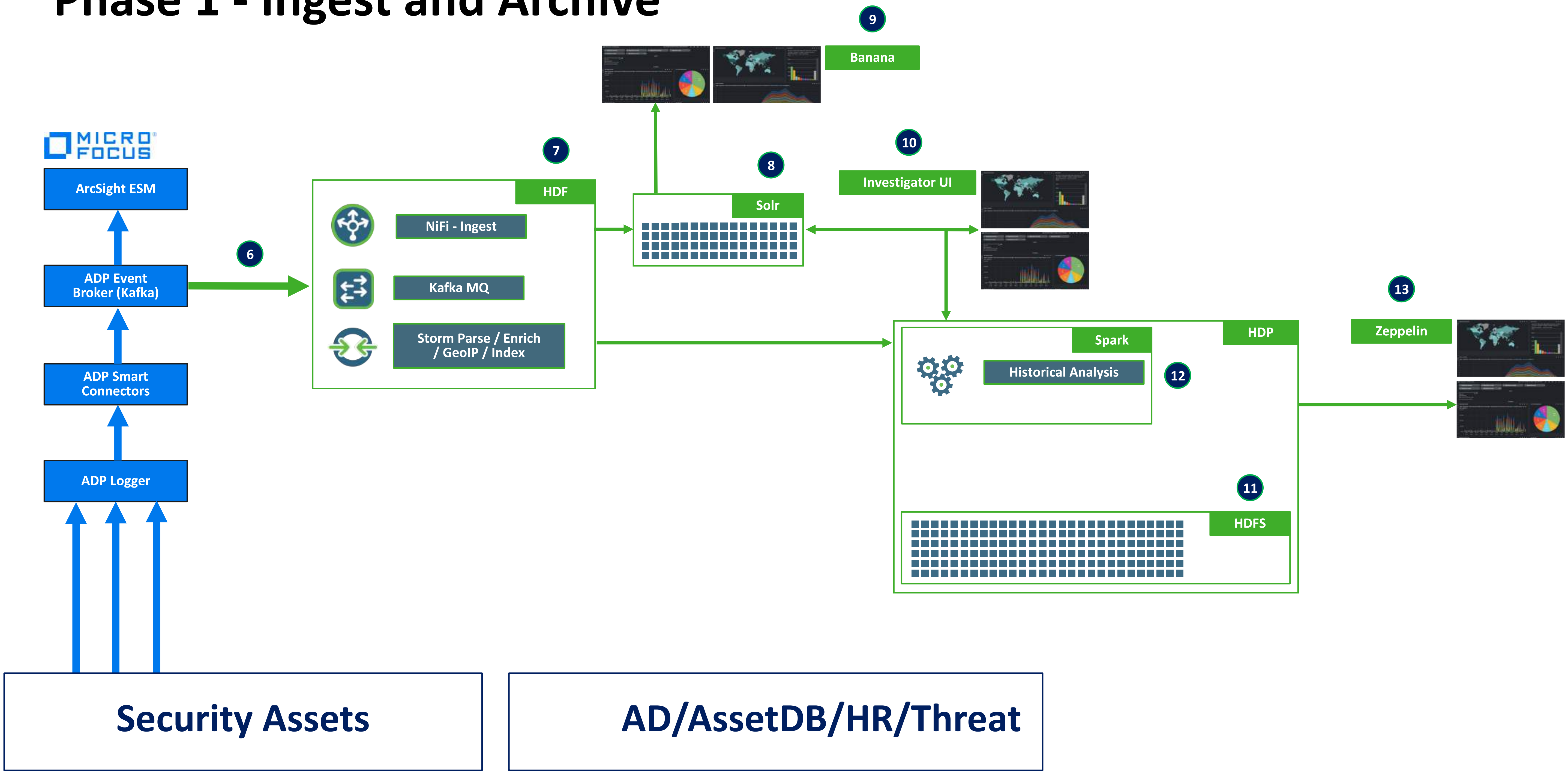
2

Security Assets

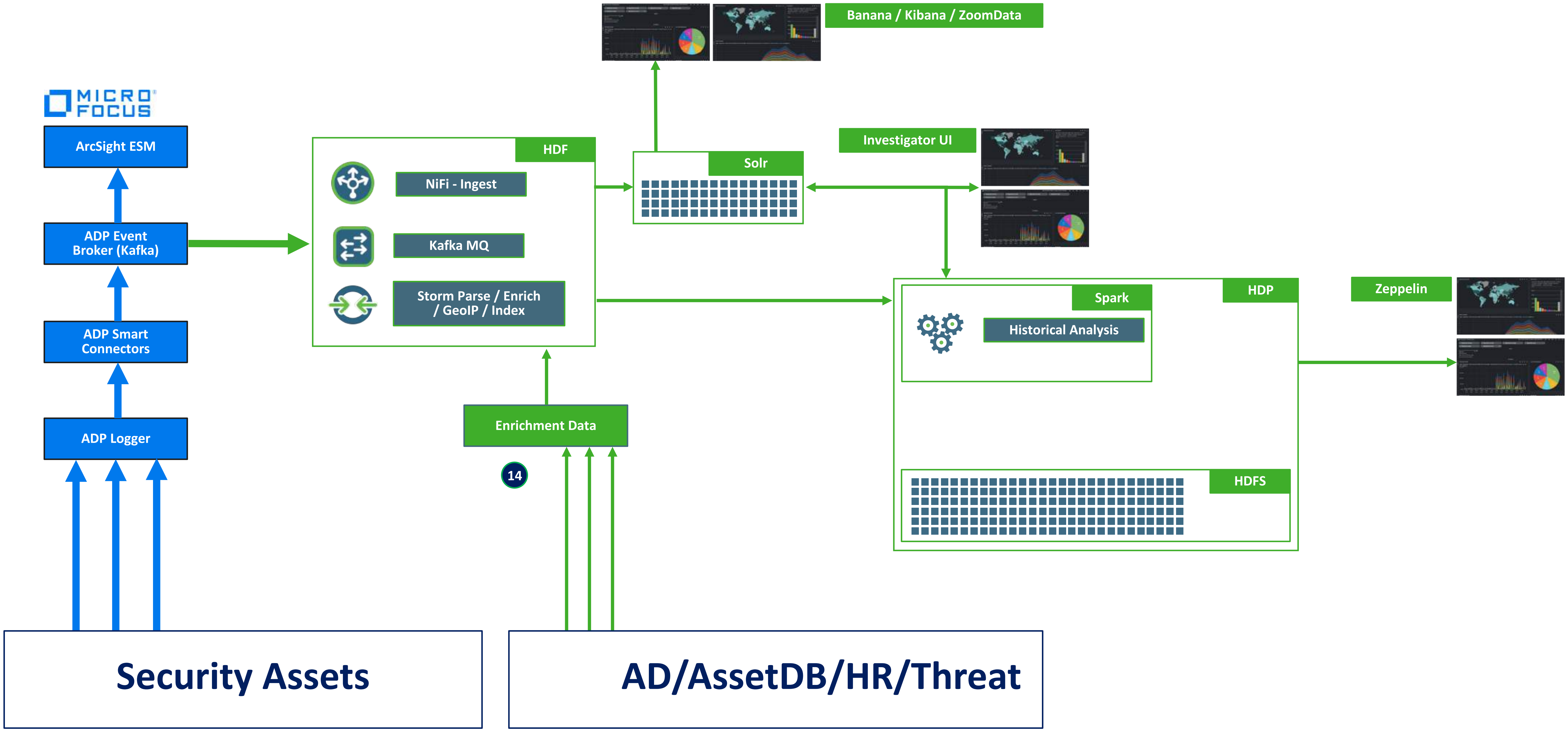
3

AD/AssetDB/HR/Threat

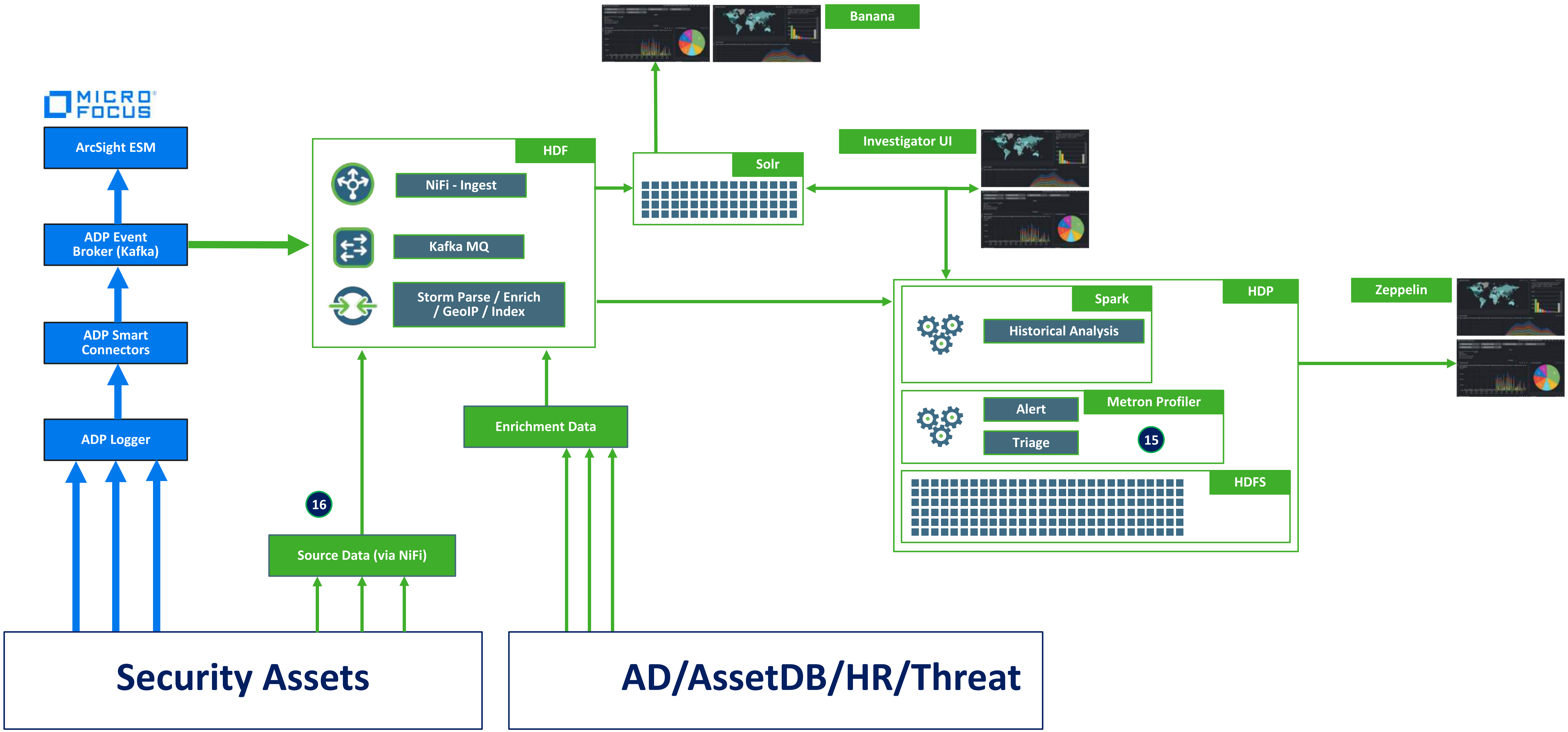
Phase 1 - Ingest and Archive



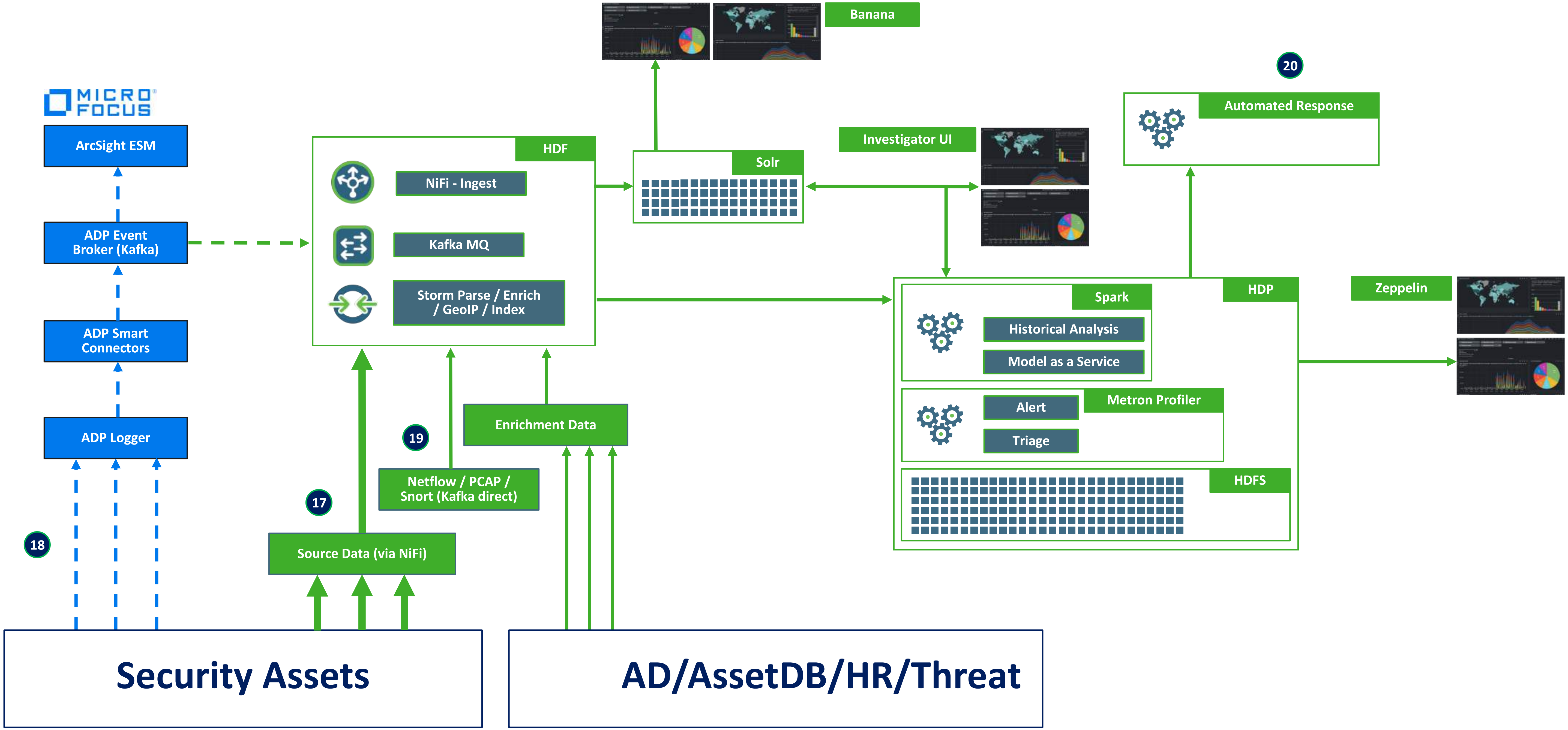
Phase 2 – Enrich and Threat Intel



Phase 3 – NiFi Data Ingestion + Analytics / UEBA Profiling



Phase 4 – ArcSight Logger Migration + New Data Sources



Considerations for Sizing Apache Metron

Sizing an HCP deployment

- Events per second (average and peak)
- Retention time for Hot / Warm / Cold zones
- Enrichments
- Node sizing
- I/O Considerations
- PCAP?

3 Months

🟥 Fast indexed layer (Solr / ES) ~3 months

🟨 Warm HDFS layer ~3 months

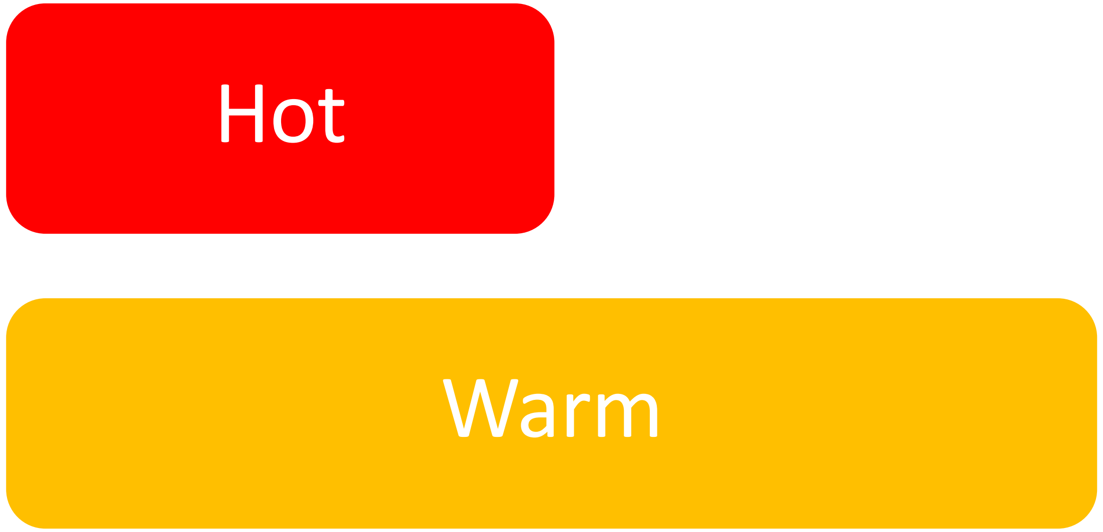
Hot

Warm

12 Months

🟥 Fast indexed layer (Solr / ES) ~3 months

🟨 Warm HDFS layer ~12 months

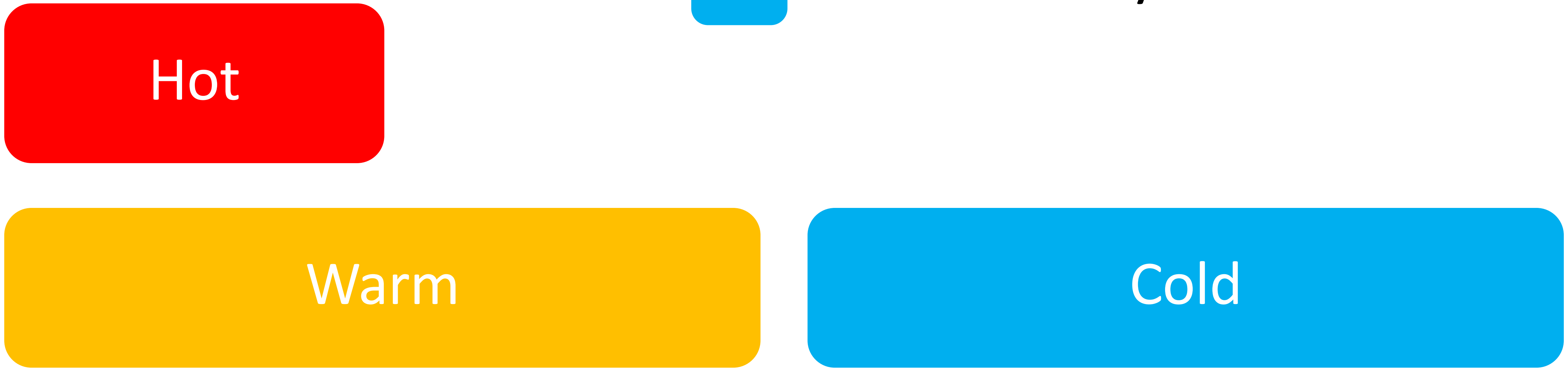


24 Months

🟥 Fast indexed layer (Solr / ES) ~3 months

🟨 Warm HDFS layer ~12 months

🟦 Cold HDFS layer +12 months



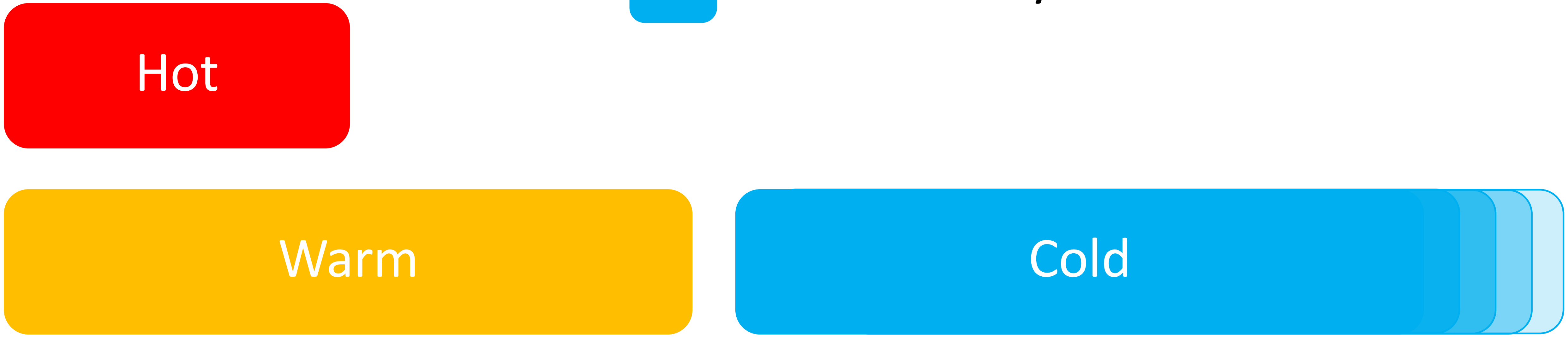
Hot

Warm

Cold

Beyond 24 months

- 🟥 Fast indexed layer (Solr / ES) ~3 months
- 🟨 Warm HDFS layer ~12 months
- 🟦 Cold HDFS layer +12 months



Hot

Warm

Cold

Questions?

- Roaring Elephant Podcast -
Bite-Sized Big Data

Appendix