

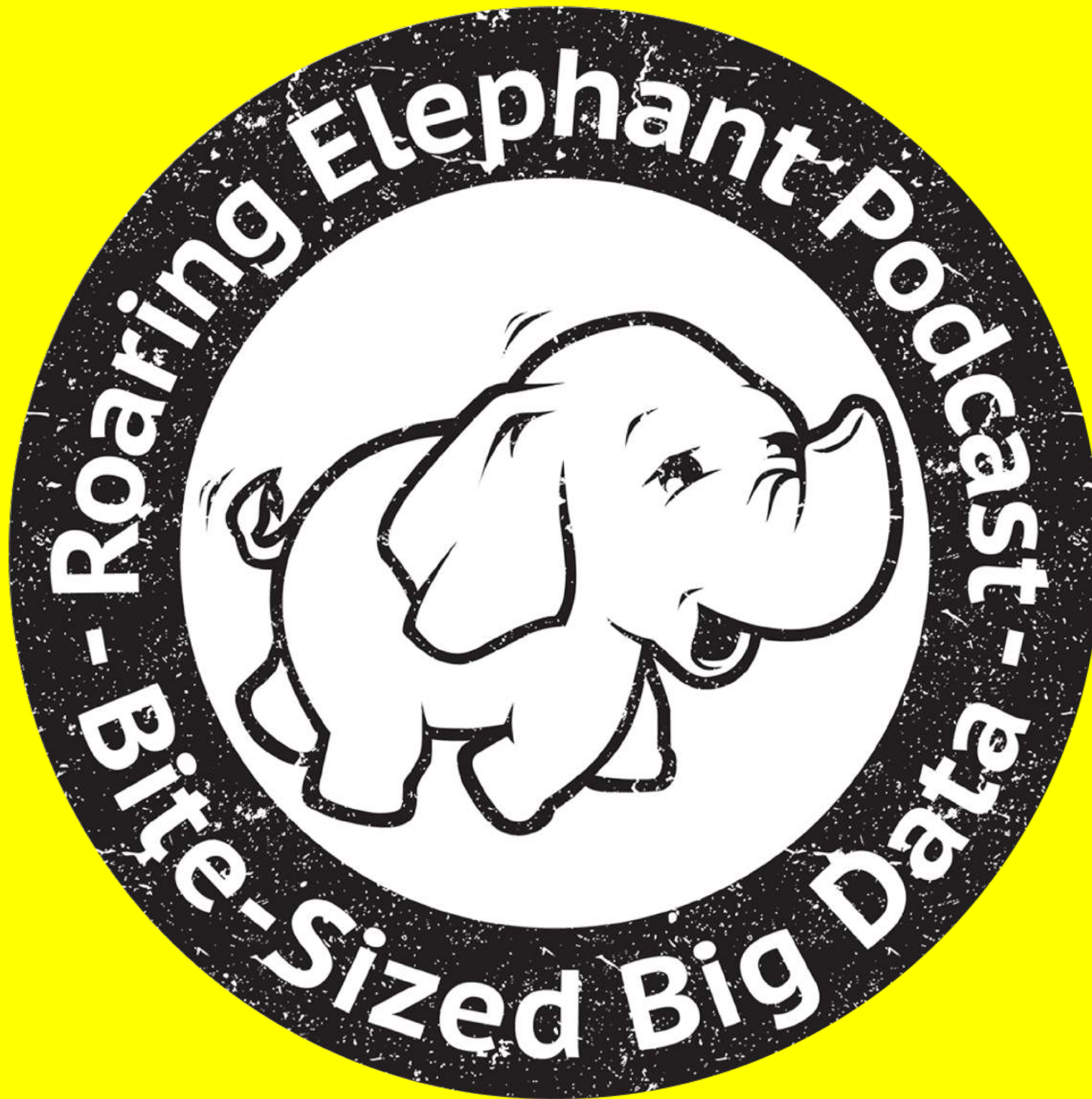
Apache Metron in the Real World

Dave Russell - Hortonworks

www.roaringelephant.org



Who am I?



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/?

35 Months



FB/Cambridge
Analytica

48 Months



THE
LINUX
FOUNDATION

**“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],
Indonesia
Hezhou 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



UNDER ARMOUR®

150M Records



92M 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

? ~ █

Searches ▾

source:type:auth

ⓧ

All time ▾





Alerts (20543)







ACTIONS ▾

Filters

enrichm...:country 0 ▾

host 0 ▾

ip_dst_addr 0 ▾

ip_src_addr 0 ▾

source:type 1 ▾

Group By

1
source:type























0
ip_dst_addr

0
host

0
enrichm...:country

0
ip_src_addr

UnGroup

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	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)	
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	

Searches ▾ `source:type:auth`

✕

All time ▾

Alerts (20543)

ACTIONS ▾

Filters

enrichm ▾

host ▾

ip_d ▾

ip_ ▾

so ▾

0 ip_dst_addr

0 host

0 enrichm...:country

0 ip_src_addr

UnGroup

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 03:57:13	auth	NEW	C612	C965	The distin...ian (1.00)	U22	
100	2017-05-25 03:59:30	auth	NEW	C467	C506	The distin...ian (1.00)	U22	
100	2017-05-25 04:03:15	auth	NEW	C612	C965	The distin...ian (1.00)	U22	
100	2017-05-25 04:03:24	auth	NEW	C467	C506	The distin...ian (1.00)	U22	
100	2017-05-25 04:03:24	auth	NEW	C625	C246	The distin...ian (1.00)	U22	
100	2017-05-25 03:56:42	auth	NEW	C528	C477	The distin...ian (1.00)	U22	
100	2017-05-25 03:56:47	auth	NEW	C61	C61	The distin...ian (1.00)	U66	
100	2017-05-25 03:58:46	auth	NEW	C528	C477	The distin...ian (1.00)	U22	
100	2017-05-25 03:59:23	auth	NEW	C612	C1143	The distin...ian (1.00)	U534	
100	2017-05-25 03:59:23	auth	NEW	C528	C477	The distin...ian (1.00)	U22	
100	2017-05-25 03:59:45	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)

Searches ▾ source:type:auth

0 enrichm...:country

All time 🔍 📄

Alerts (20543)

- Filters
- enrichm...:country 0 ▾
 - host 0 ▾
 - ip_dst_addr 0 ▾
 - ip_src_addr 0 ▾
 - source:type 1 ▾

Group By 1 source:type 0 ip_dst_addr

Score ▾	timestamp ▾	source:type ▾	alert	threat:...0:reason ▾	user ▾	ip_src_addr	ip_dst_addr
100	2017-05-25 03:56:55	auth	NEW	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)	U22		
100	2017-05-25 03:56:56	auth	NEW	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)	U22		
100	2017-05-25 03:57:00	auth	NEW	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)	U22		
100	2017-05-25 04:03:15	auth	NEW				
100	2017-05-25 04:08:09	auth	NEW				
100	2017-05-25 03:56:55	auth	NEW				
100	2017-05-25 03:57:00	auth	NEW				
100	2017-05-25 03:57:36	auth	NEW				
100	2017-05-25 03:59:30	auth	NEW				
100	2017-05-25 04:08:24	auth	NEW				
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	The distinct number of machines that user U534 attempted to login to (2) is more than 5 standard deviations (0.29) from the median (1.00)	U534		
10	2017-05-25 03:56:47	auth	NEW	C625 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)	U22		
10	2017-05-25 03:58:46	auth	NEW	C625 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)	U22		
10	2017-05-25 03:59:23	auth	NEW	C553 C1968 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)	U22		
10	2017-05-25 03:59:23	auth	NEW	C523 C1968 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)	U22		
10	2017-05-25 03:59:45	auth	NEW	C467 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)	U22		

What is Apache Metron?

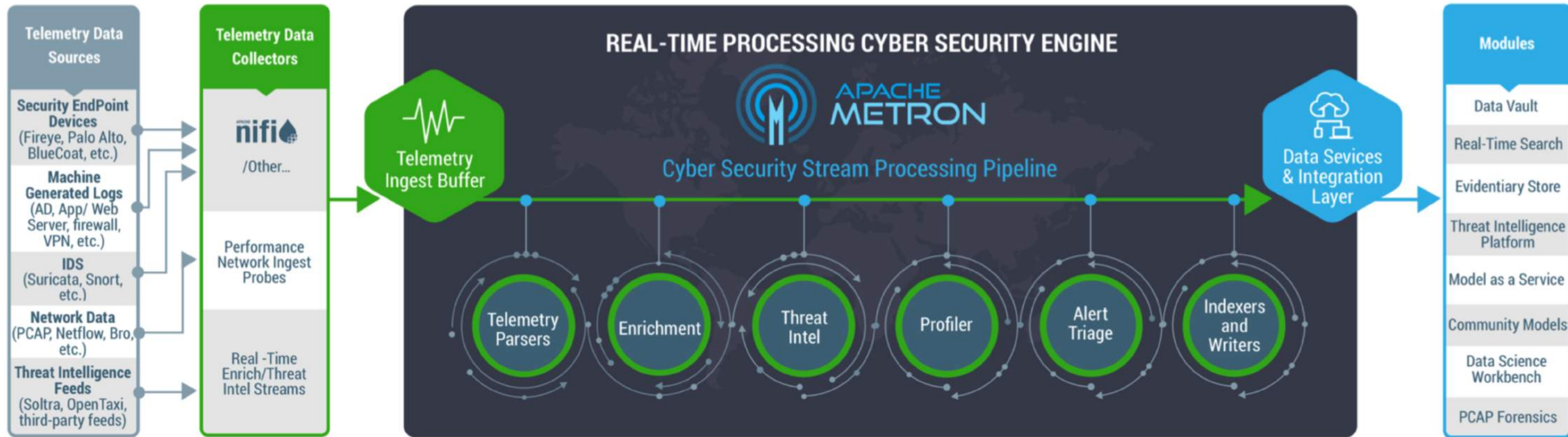
Built on top on proven open source big data technology



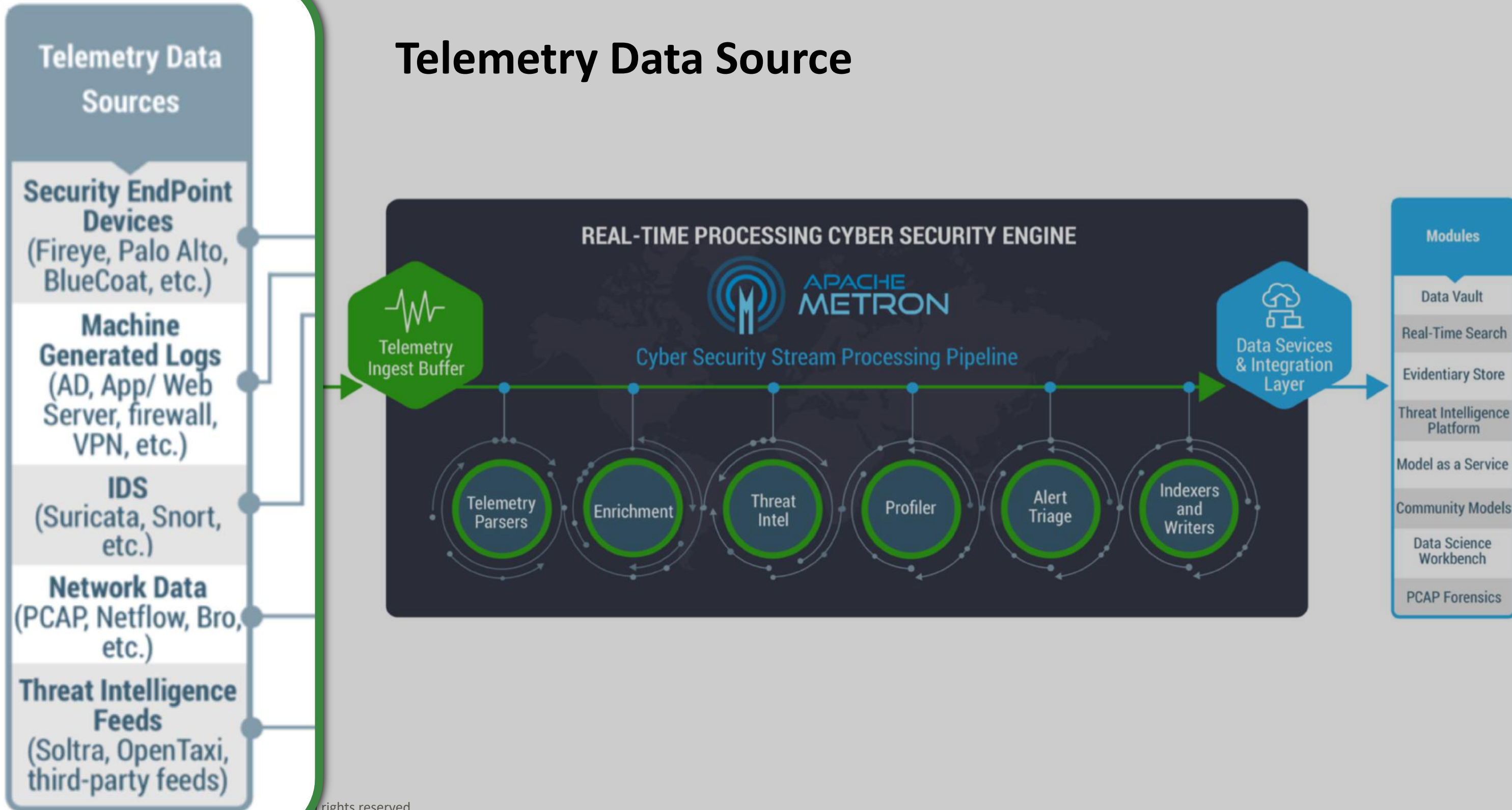
Ambari



An architecture for real-time cybersecurity analytics



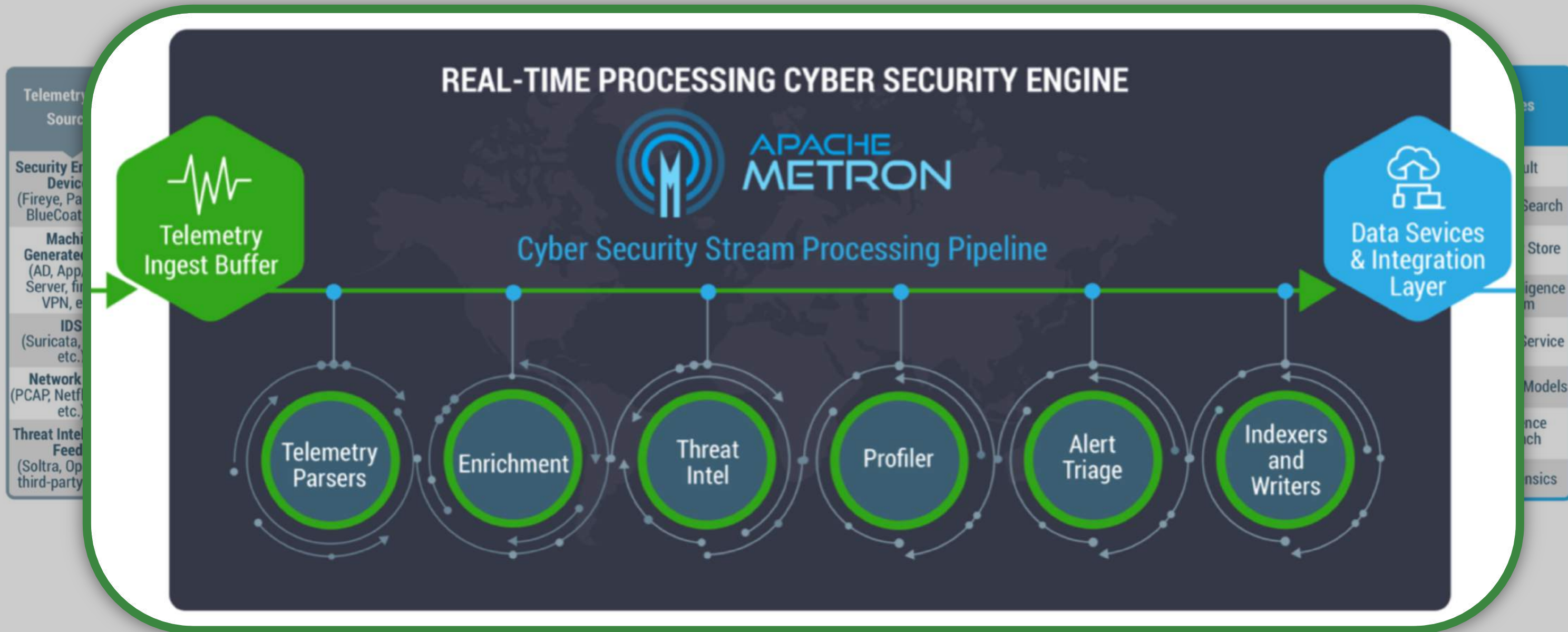
Telemetry Data Source



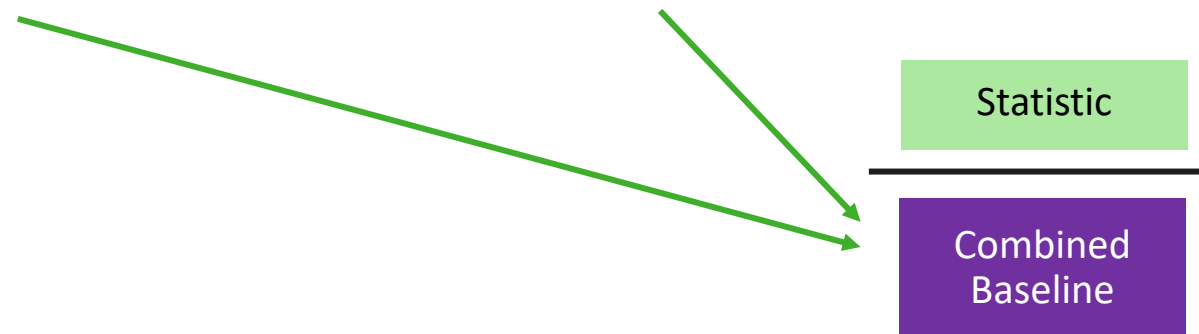
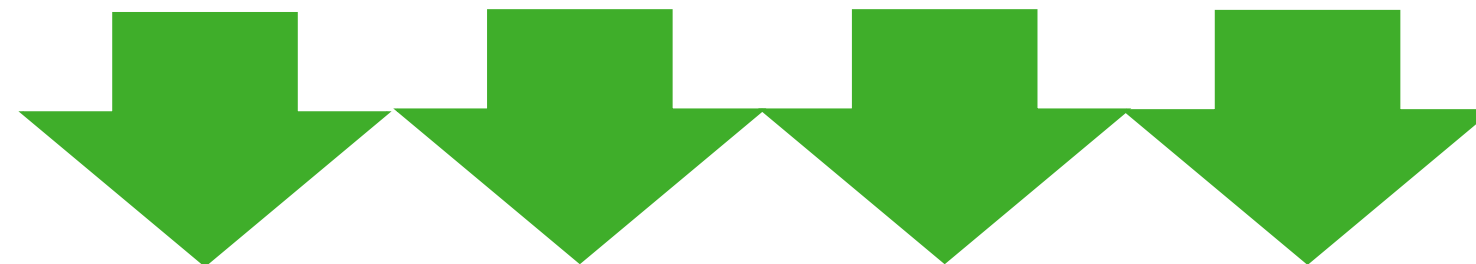
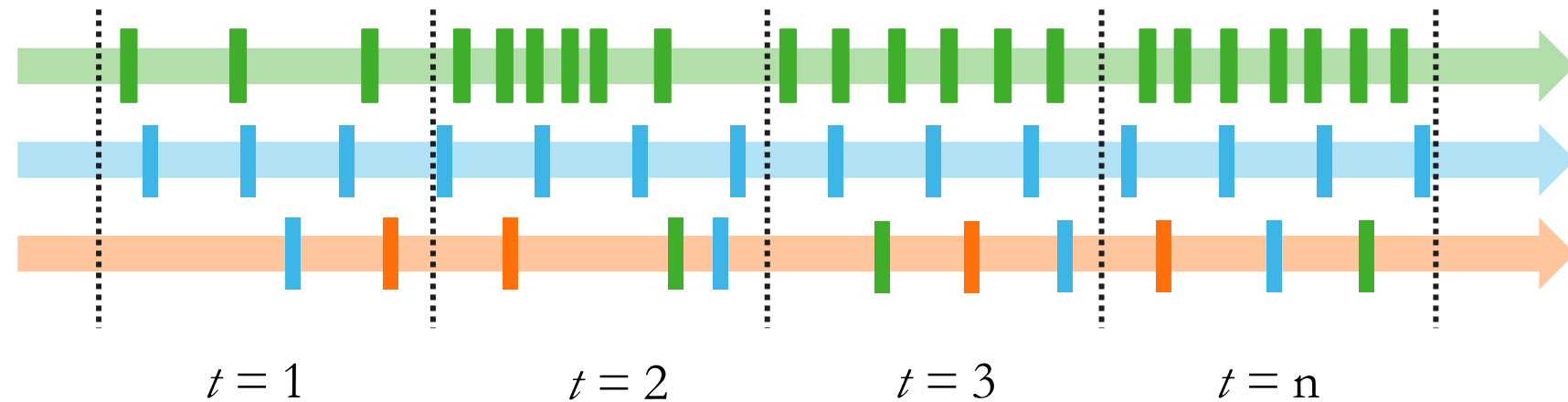
Telemetry Data Collectors



Cyber Security Stream Processing Pipeline



Profiling by time

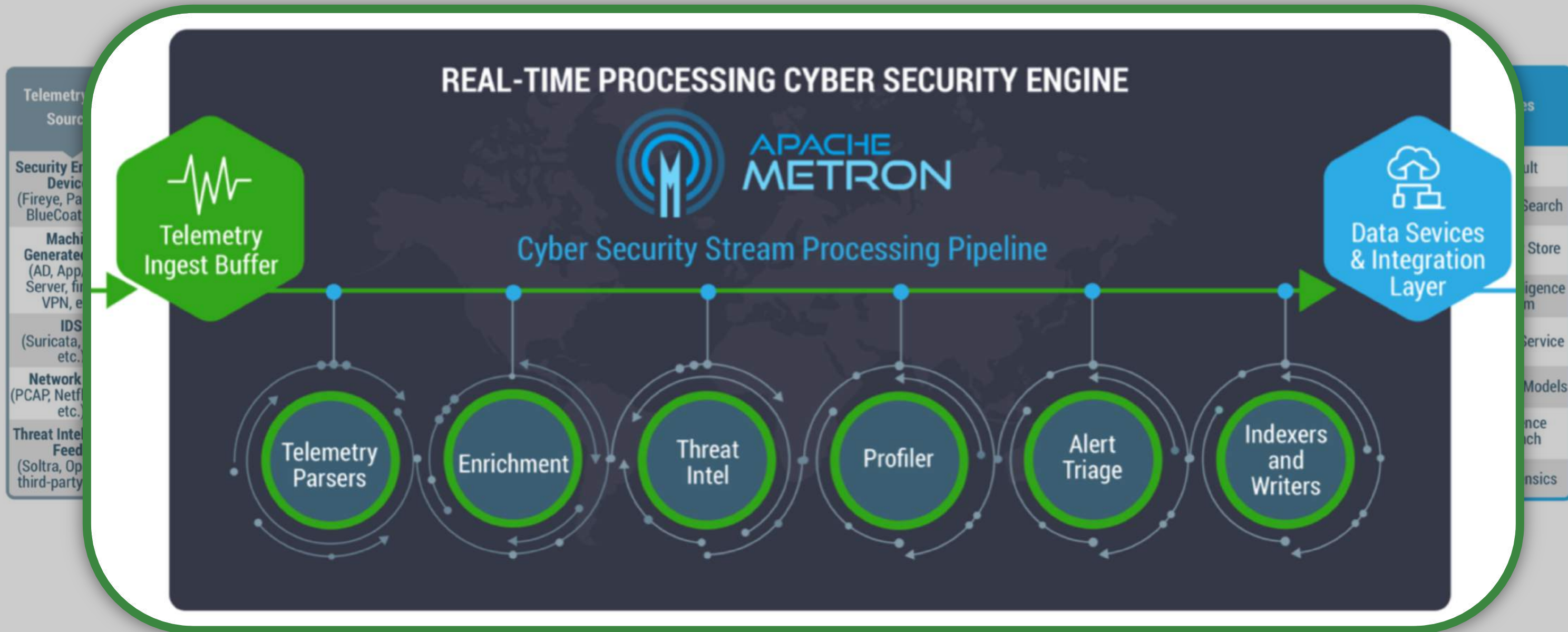


Wide range of algorithms including:

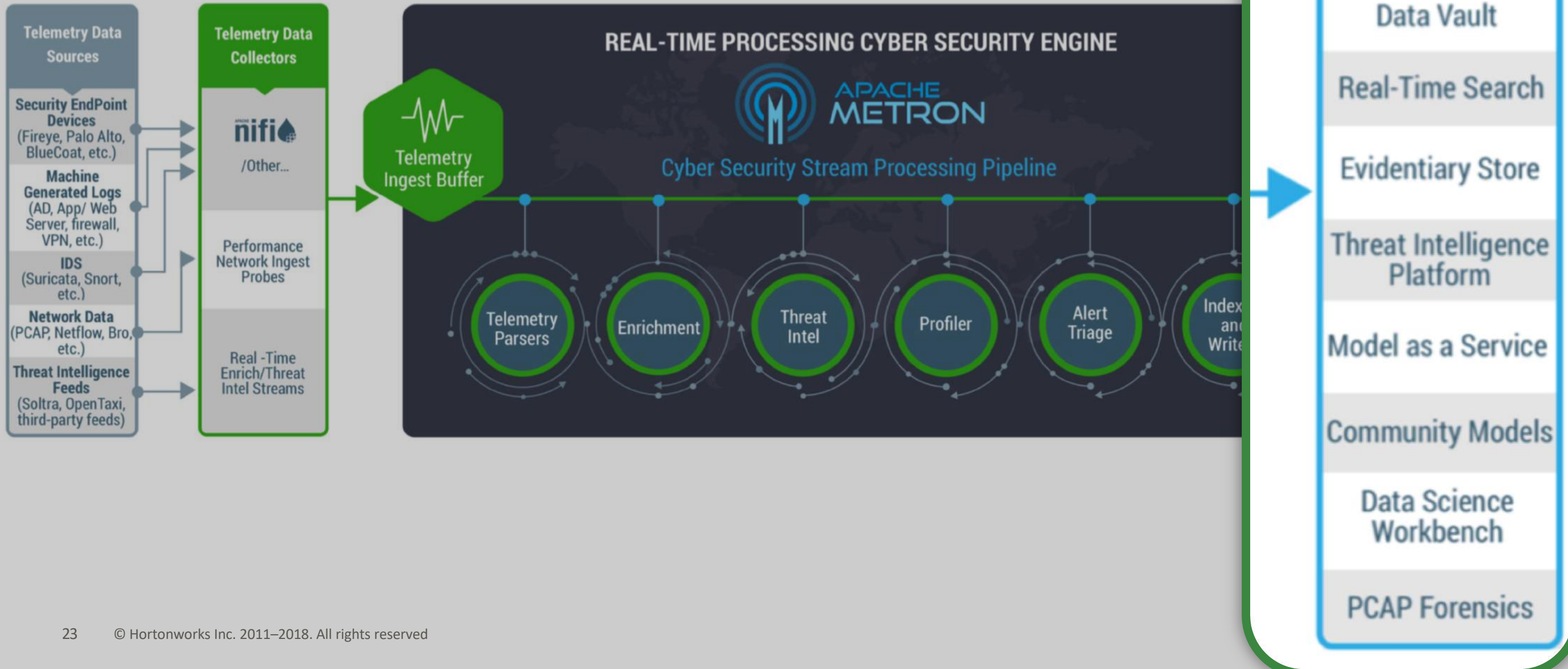
- ◆ HyperLogLogPlus
- ◆ Bloom filters
- ◆ T-digests
- ◆ Statistical Baseline
- ◆ Hashing functions
- ◆ Outlier detection
- ◆ GeoHashing over time
- ◆ Locality Sensitive Hashing



Cyber Security Stream Processing Pipeline



Apache Metron Modules



Who is Using Apache Metron (Part 1)



Capital One



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





Q Sight IT®

KPN neemt QSight IT over

[Lees het hele persbericht](#)

Oplossingen



Branches



Diensten

The Zscaler logo, consisting of a blue circular icon with a white 'Z' shape inside, followed by the word 'zscaler' in a lowercase sans-serif font.The Commvault logo, featuring a stylized 'C' icon made of three geometric shapes (a triangle, a square, and a circle) in black, white, and grey, followed by the word 'COMMVULT' in a bold, uppercase sans-serif font.The Qualys logo, featuring a red shield icon with a white 'Q' inside, followed by the word 'QUALYS' in a bold, uppercase sans-serif font.The Akamai NetSession logo, featuring the word 'Akamai' in a stylized font with a blue and orange gradient, followed by 'NETSESSION' in a smaller, uppercase sans-serif font.The Cisco logo, featuring a stylized bridge icon made of seven vertical bars of increasing height, followed by the word 'CISCO' in a bold, uppercase sans-serif font.



The Wider Apache Metron Ecosystem



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



PSSC Labs

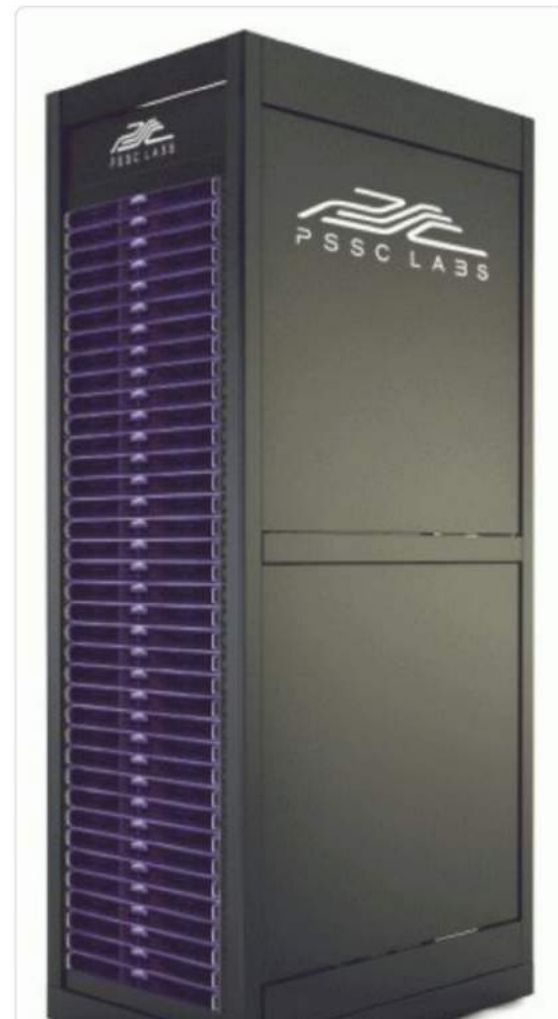
@PSSCLabs

Follow

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

#HPC #CyberSecurity #supercomputer
#bigdata #PSSCLabs

[prweb.com/releases/2018/ ...](http://prweb.com/releases/2018/)



Demo / Video

Who is Using Apache Metron (Part 2)



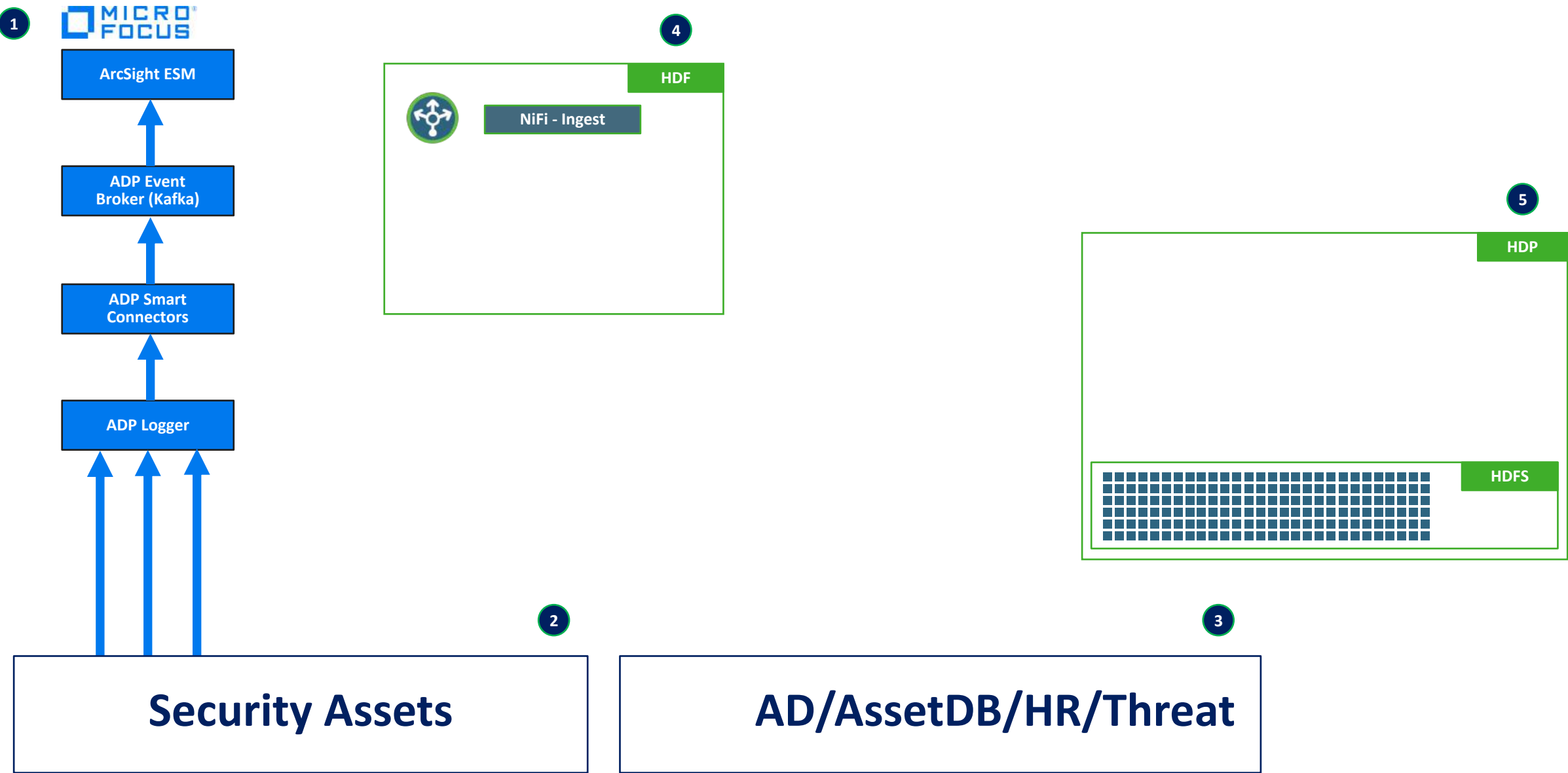




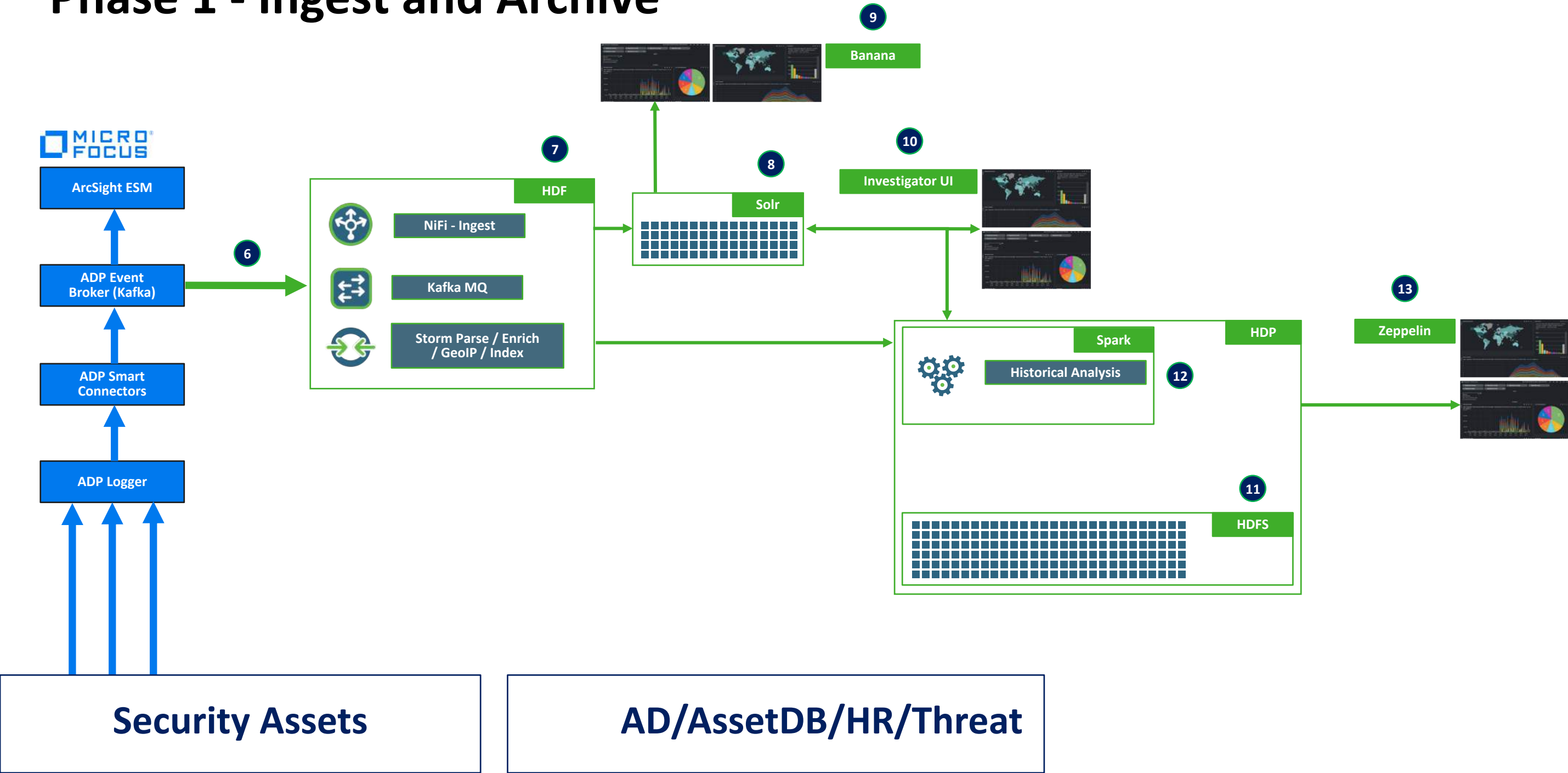


Deploying Apache Metron

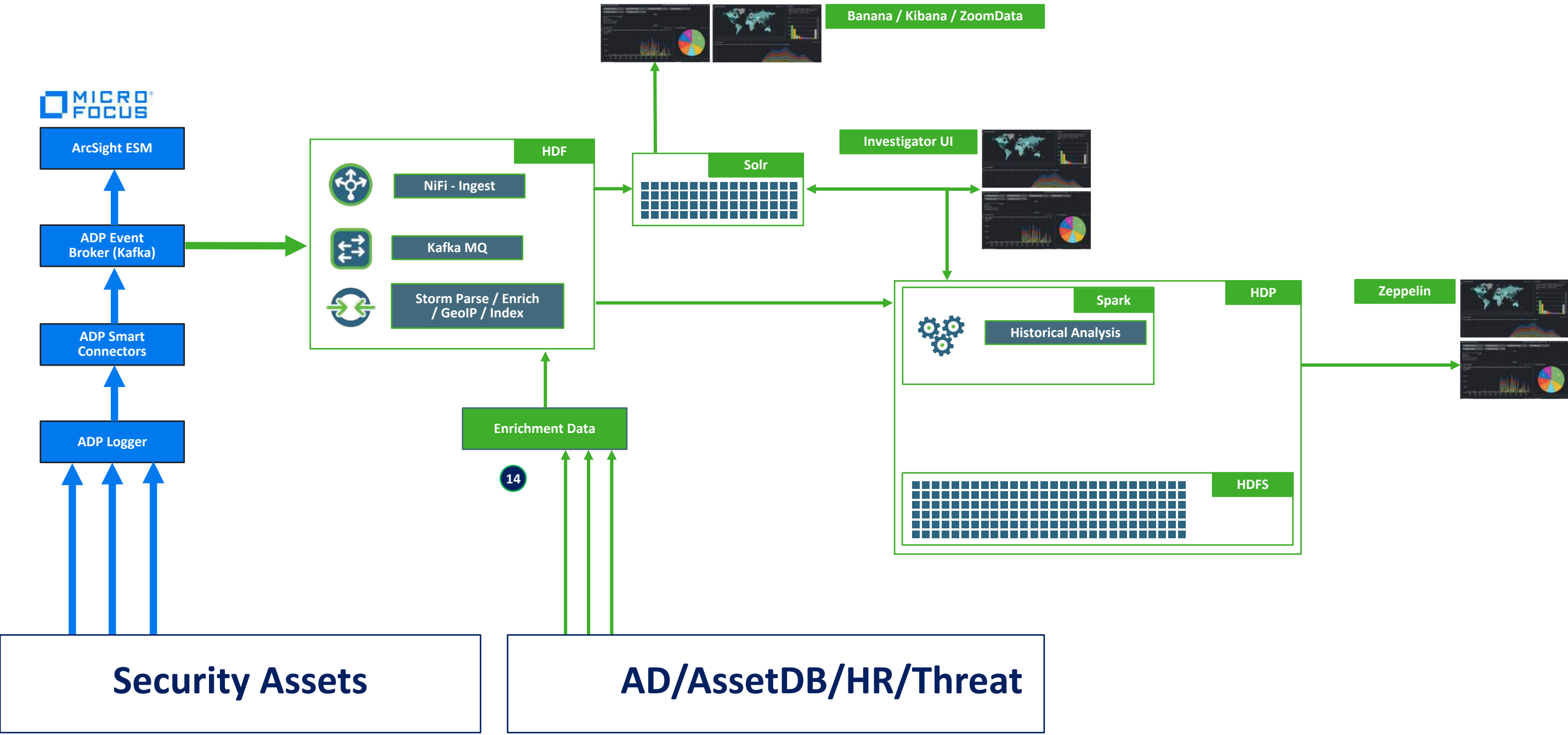
Phase 0 – Current State



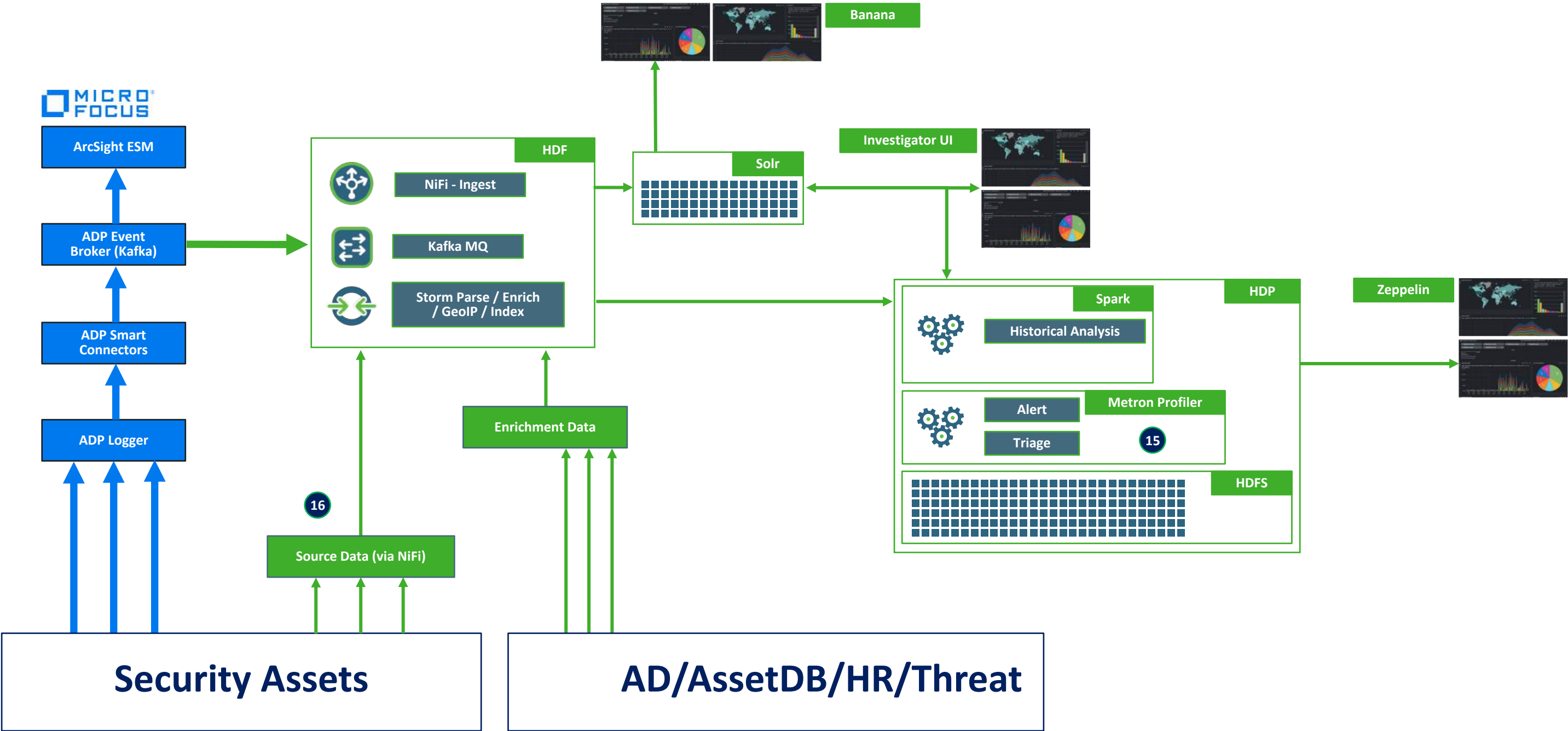
Phase 1 - Ingest and Archive



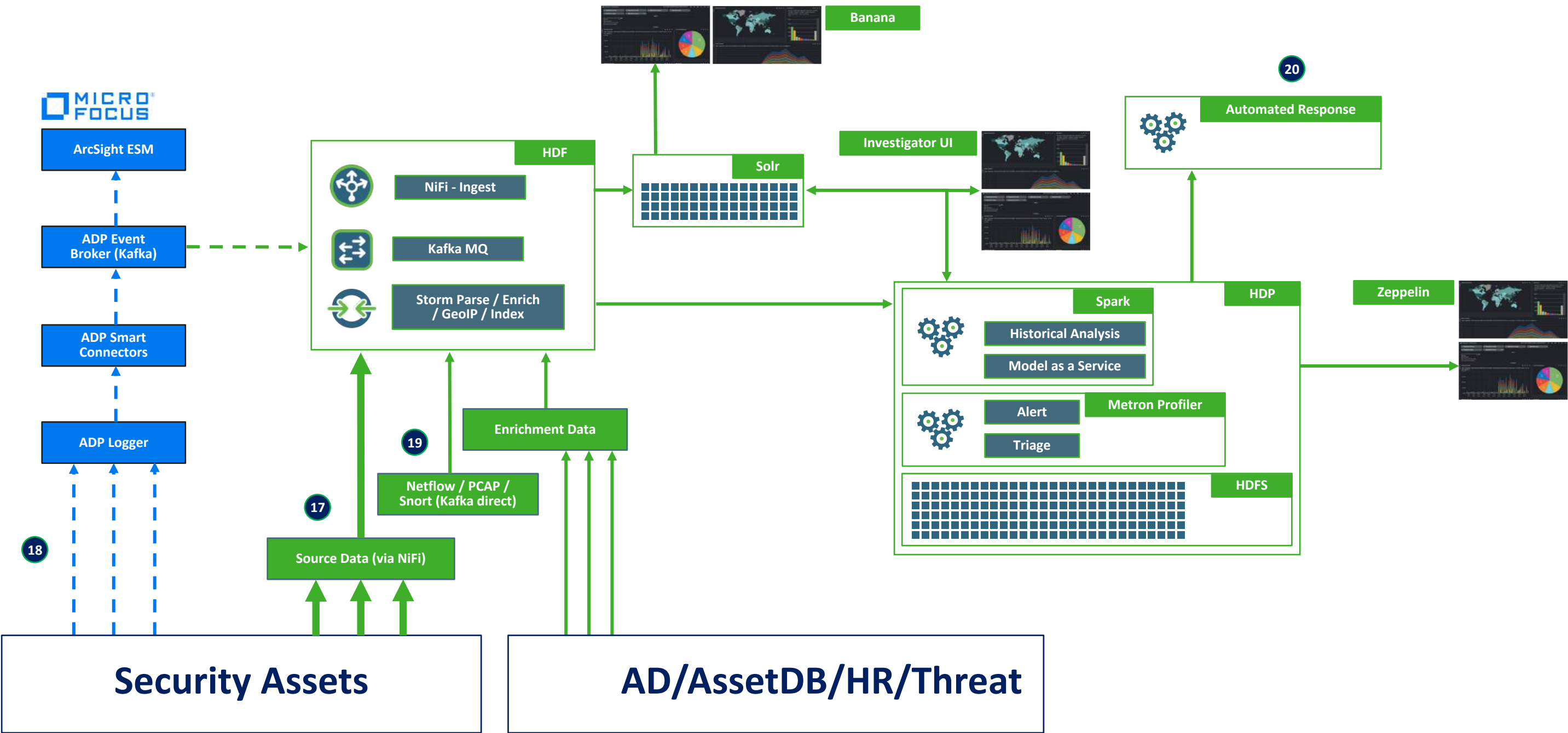
Phase 2 – Enrich and Threat Intel



Phase 3 – NiFi Data Ingestion + Analytics / UEBA Profiling



Phase 4 – ArcSight Logger Migration + New Data Sources





Questions?



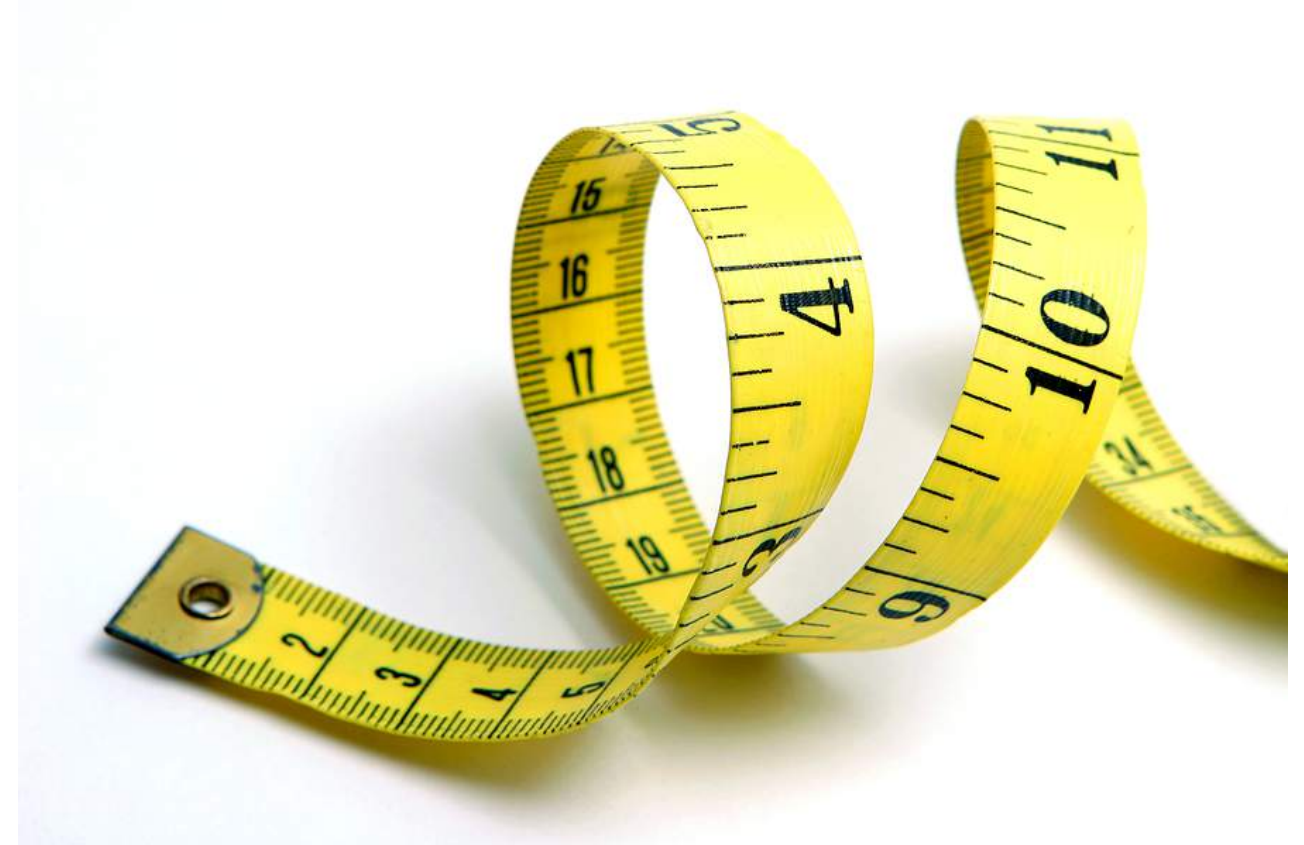
Appendix



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

Hot

Warm

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

Apache Metron Data Sheet

- Ingest:
 - Apache NiFi: syslog, socket, file, web services, SQL, RDBMS, Windows Event Log, FTP, MQ, JMS, Splunk and others
 - High-performance DPDK Packet Capture
- Parsers:
 - Cisco ASA
 - Bluecoat
 - Fireeye
 - Palo Alto
 - SourceFire
 - WebSphere
 - Snort IDS
 - Bro DPI
 - Netflow, IPFIX
 - Grok (Custom)
- Java (Custom)
- JSON
- CEF, LEEF (ArcSight, Qradar compat.)
- Applications: DHCPD, AD
- Enrichments and threat feeds:
 - Geo
 - Whois
 - HBase
 - JDBC
 - Stellar
 - CSV
 - Stix, Taxii threat intel feeds
- Analytics features:
 - Profiler and statistical baselining engine
 - Model Services for advanced ML
- Threat Triage rules and scoring engine
- Indexing and search:
 - Elasticsearch, Kibana
 - Solr
 - HDFS
 - Kafka
- Data science features:
 - Spark Machine Learning
 - Zeppelin notebooks and reporting
 - Wide partner eco-system
- Forensic features:
 - PCAP inspector
 - PCAP query
 - Long term data store