2nd CISO 360 CONGRESS 2018

INTELLIGENCE | INNOVATION | INSPIRATION

PULSE CONFERENCES

Case study
Open-source stack of tools for strategic security data analytics

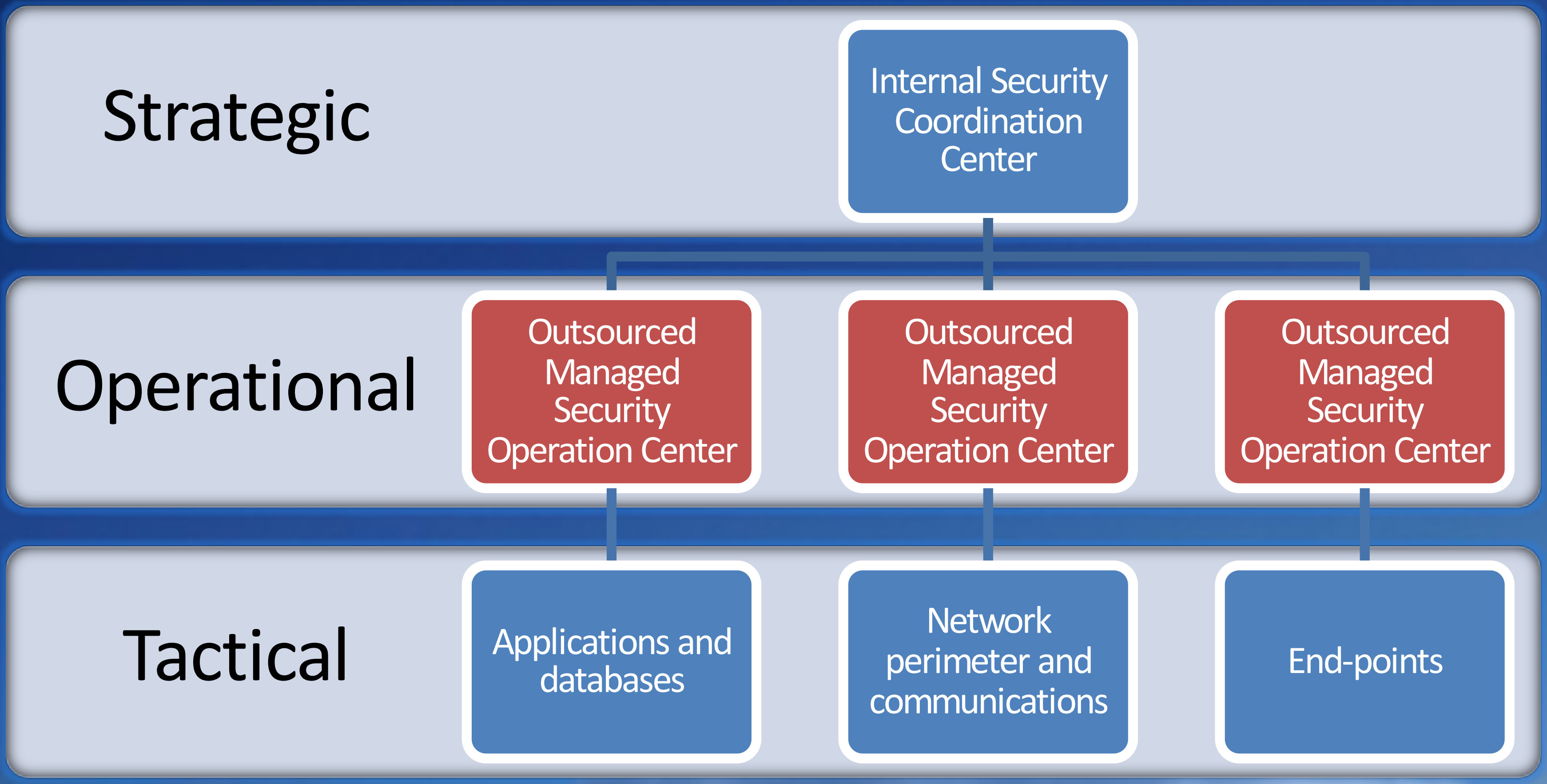
Dr. Viktor Polic

CISO, International Labour Organization

Adjunct Professor of Computer Science,
Webster University Geneva

The views expressed herein are those of the author and do not necessarily reflect the views of the International Labour Organization.

ISMS decision levels and risk intelligence



Strategic

Internal Security Coordination Center

Operational

Outsourced Managed Security Operation Center

Outsourced Managed Security Operation Center

Outsourced Managed Security Operation Center

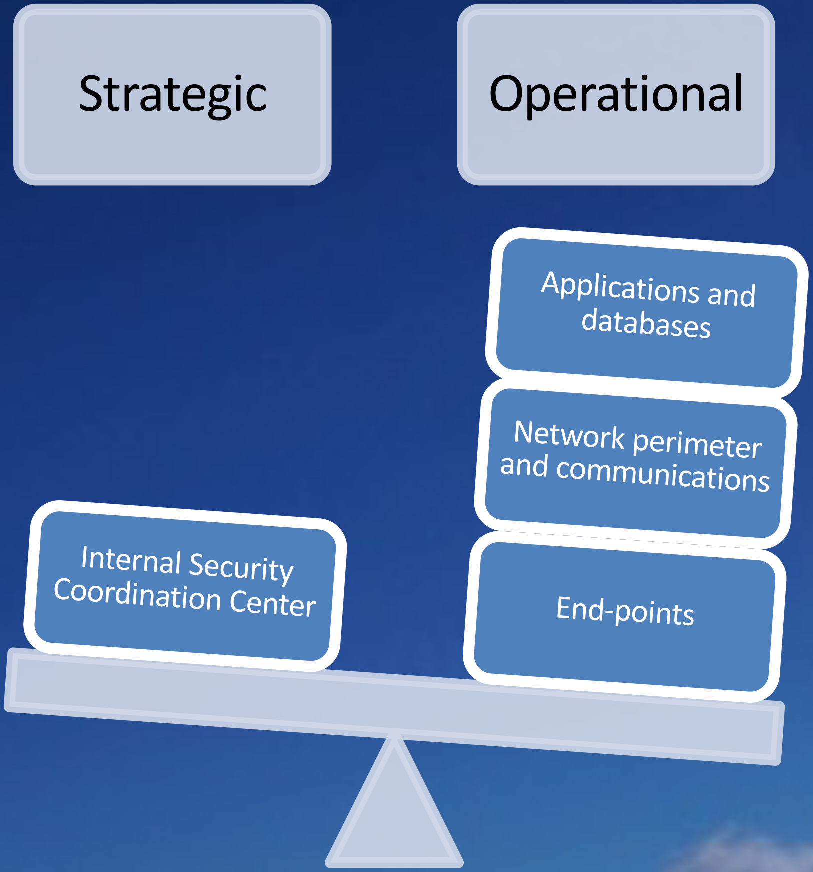
Tactical

Applications and databases

Network perimeter and communications

End-points

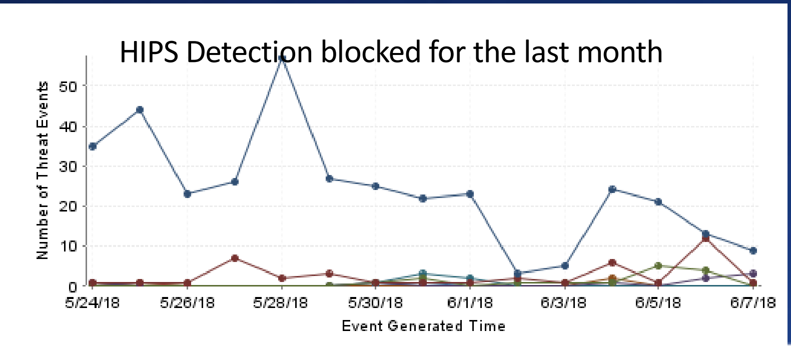
ISMS decision levels data loads

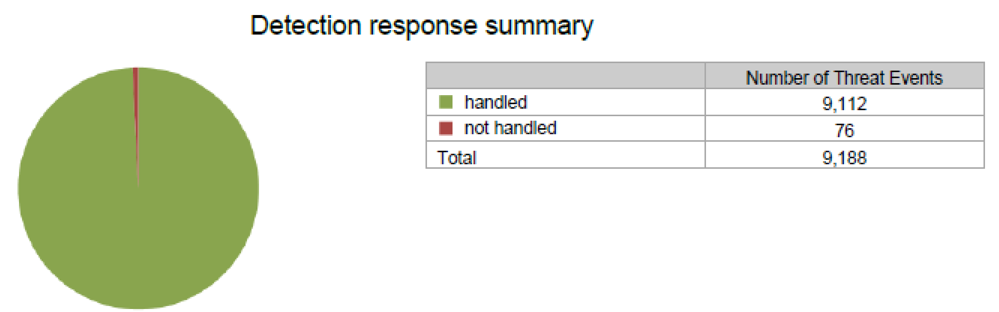

Strategic
Operational
Applications and databases
Network perimeter and communications
End-points
Internal Security Coordination Center

Operational data loads are higher in volumes and velocity because of log retention requirements and amount of signals before data enrichment.

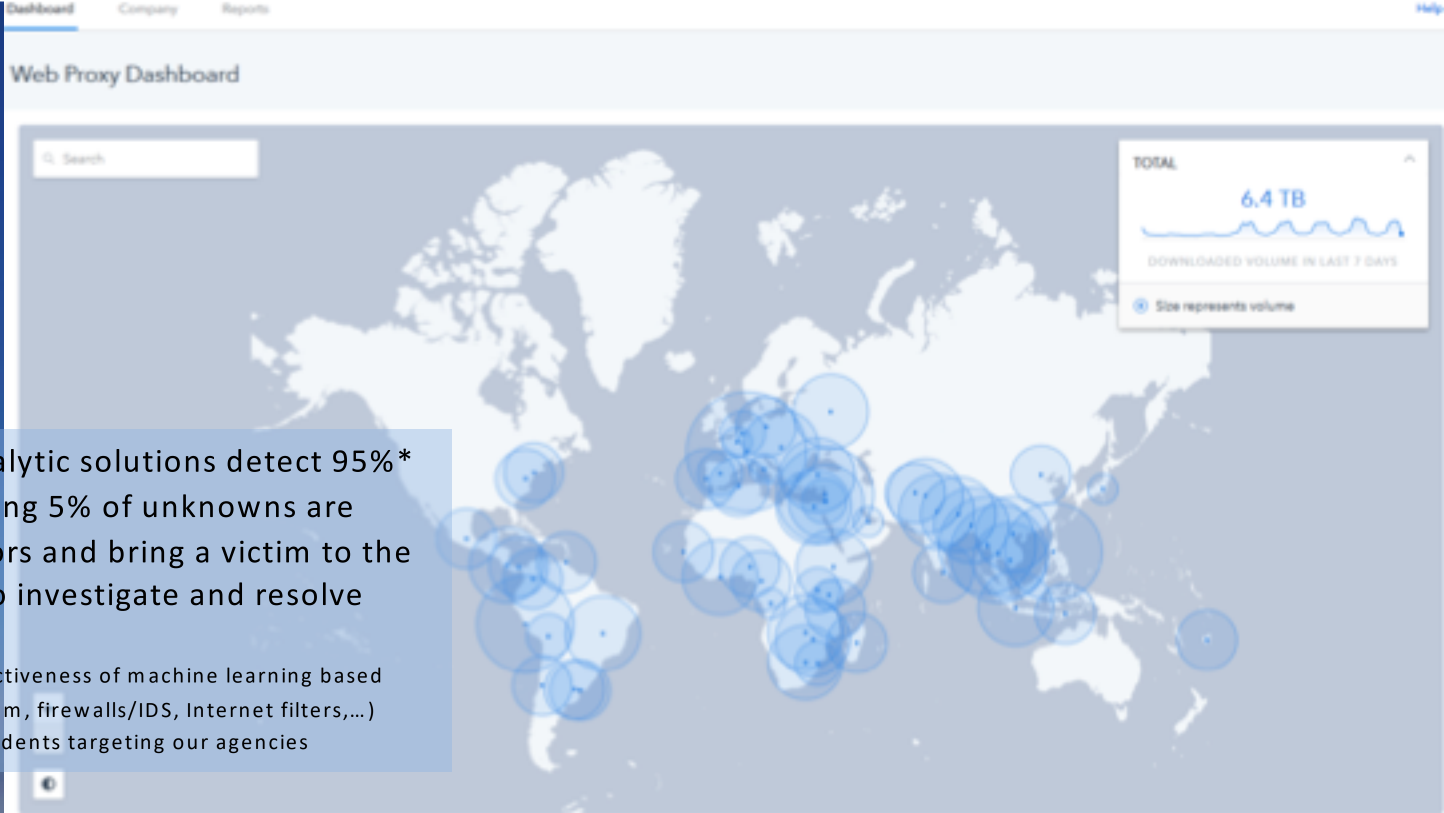
Data analytics automation is more effective within operational SoCs since data signals are grouped by technology.

Managed SOC challenges




Detection response summary

	Number of Threat Events
🟩 handled	9,112
🟥 not handled	76
Total	9,188

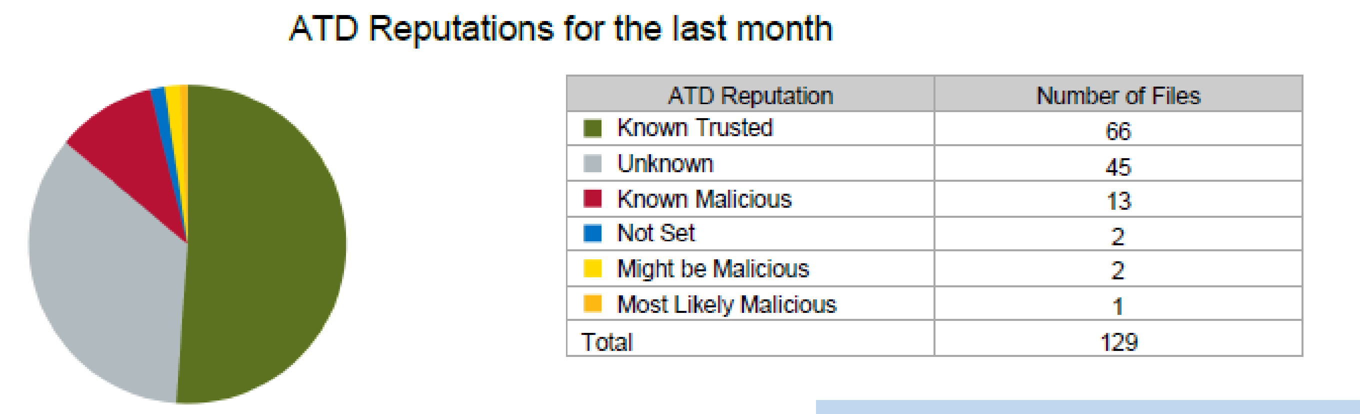


Commercial security data analytic solutions detect 95%* of known threats but remaining 5% of unknowns are exploited by determined actors and bring a victim to the headlines and cost millions to investigate and resolve (and take years!)**

*personal liberal estimate based on effectiveness of machine learning based security technologies (anti-virus, anti-spam, firewalls/IDS, Internet filters,…)

** based on post-mortem analysis of incidents targeting our agencies

Risk intelligence challenges and solutions



ATD Reputations for the last month

ATD Reputation	Number of Files
🟩 Known Trusted	66
⬜ Unknown	45
🟥 Known Malicious	13
🟦 Not Set	2
🟨 Might be Malicious	2
🟧 Most Likely Malicious	1
Total	129

Challenges

- Timeliness of threat intelligence
- Quality of intelligence data
- Usability of data (taxonomies)
- Relevance to organizational risk context

Leveraging the state of the art technology such as machine learning to resolve business problems (to manage innovation and opportunities requires using it also to manage related risks!

Solutions

- Decision support based on factual data driven modelling
- Unknown unknowns (unforeseen and unexpected events) derived from threat intelligence supported data repositories (i.e. data enriched security logs)
- To be proactive, batch processing is gradually replaced by data streaming
- Open architecture and open data (with information security in mind!) is strategically critical for internal and external collaboration

On-premise Security Coordination System Infrastructure

Apache Solr

APACHE METRON

kafka
A distributed streaming platform

STORM

elastic

hadoop YARN

FLUME

APACHE HBASE

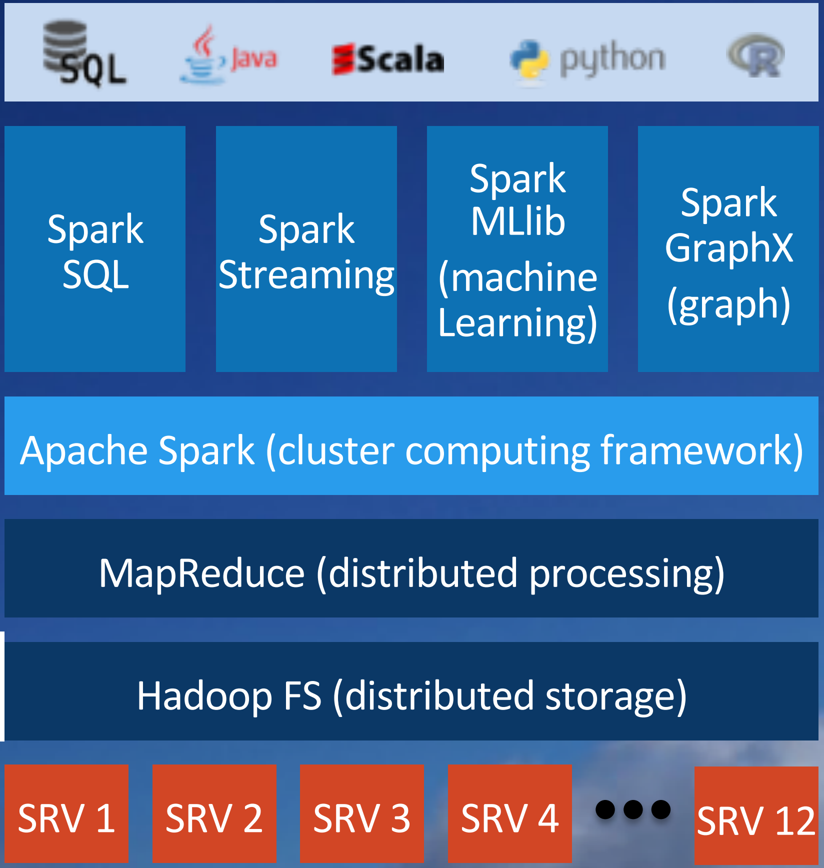
HIVE

nifi

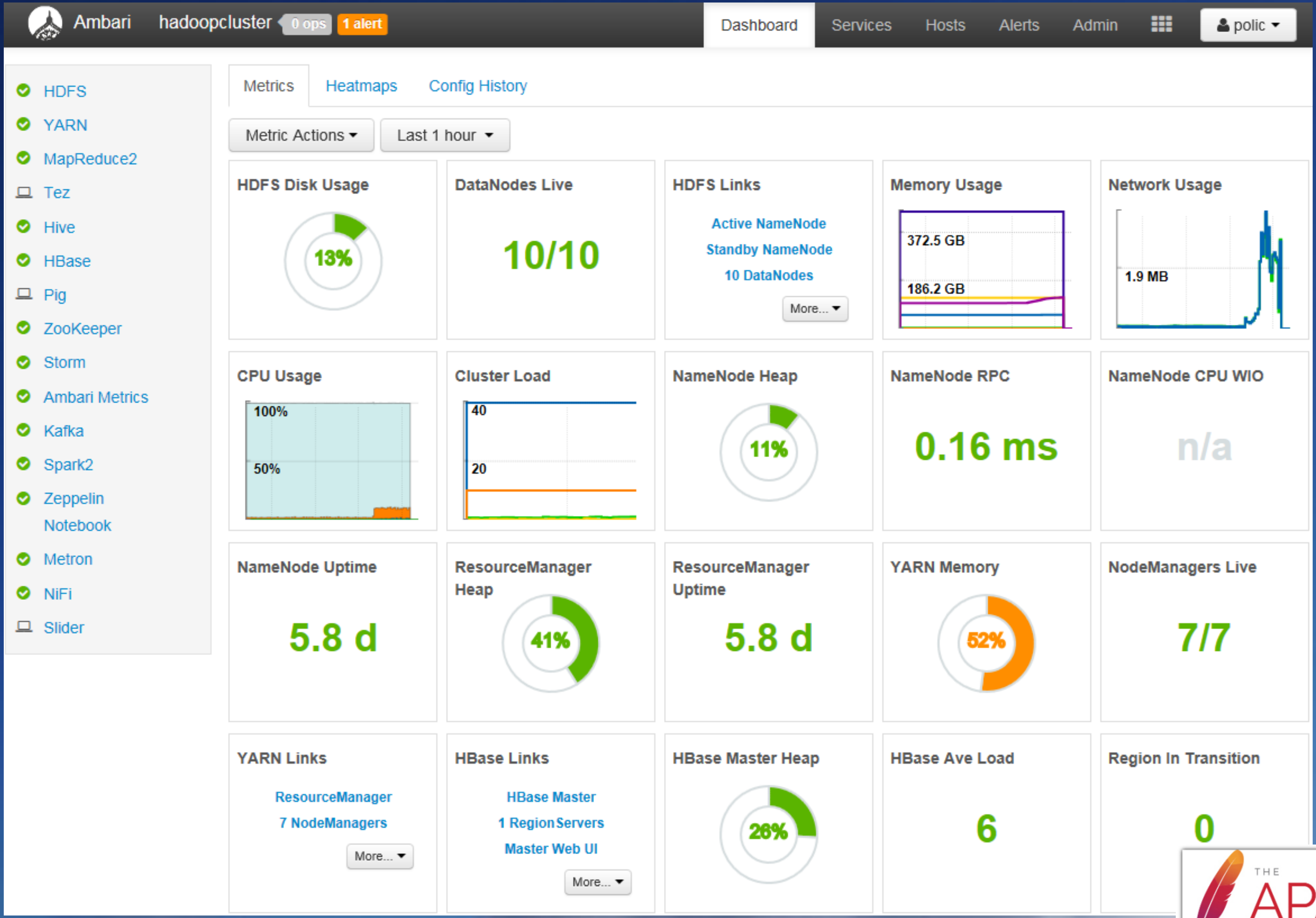
Spark

hadoop

Apache Ambari


SQL
Java
Scala
python
R
Spark SQL
Spark Streaming
Spark MLlib (machine Learning)
Spark GraphX (graph)
Apache Spark (cluster computing framework)
MapReduce (distributed processing)
Hadoop FS (distributed storage)
SRV 1
SRV 2
SRV 3
SRV 4
•••
SRV 12

100% open-source based



HDFS Disk Usage: 13%

DataNodes Live: 10/10

HDFS Links: Active NameNode, Standby NameNode, 10 DataNodes

Memory Usage: 372.5 GB, 186.2 GB

Network Usage: 1.9 MB

CPU Usage: 100%, 50%

Cluster Load: 40, 20

NameNode Heap: 11%

NameNode RPC: 0.16 ms

NameNode CPU WIO: n/a

NameNode Uptime: 5.8 d

ResourceManager Heap: 41%

ResourceManager Uptime: 5.8 d

YARN Memory: 52%

NodeManagers Live: 7/7

YARN Links: ResourceManager, 7 NodeManagers

HBase Links: HBase Master, 1 RegionServers, Master Web UI

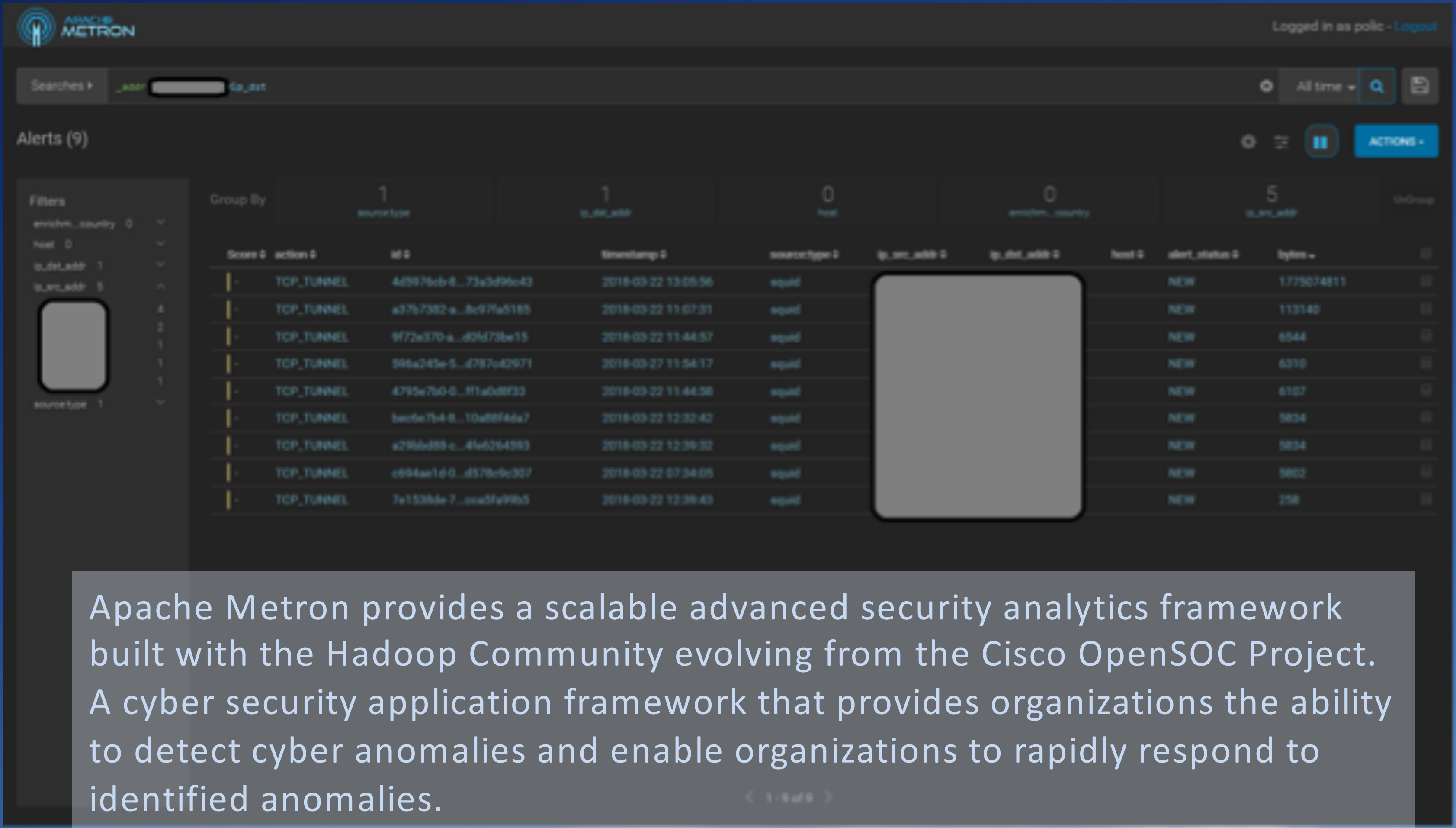
HBase Master Heap: 26%

HBase Ave Load: 6

Region In Transition: 0

THE APACHE SOFTWARE FOUNDATION
@ 19
COMMUNITY-LED DEVELOPMENT
The Apache Way

Open SoC and more…



_addr [redacted] &ip_dst

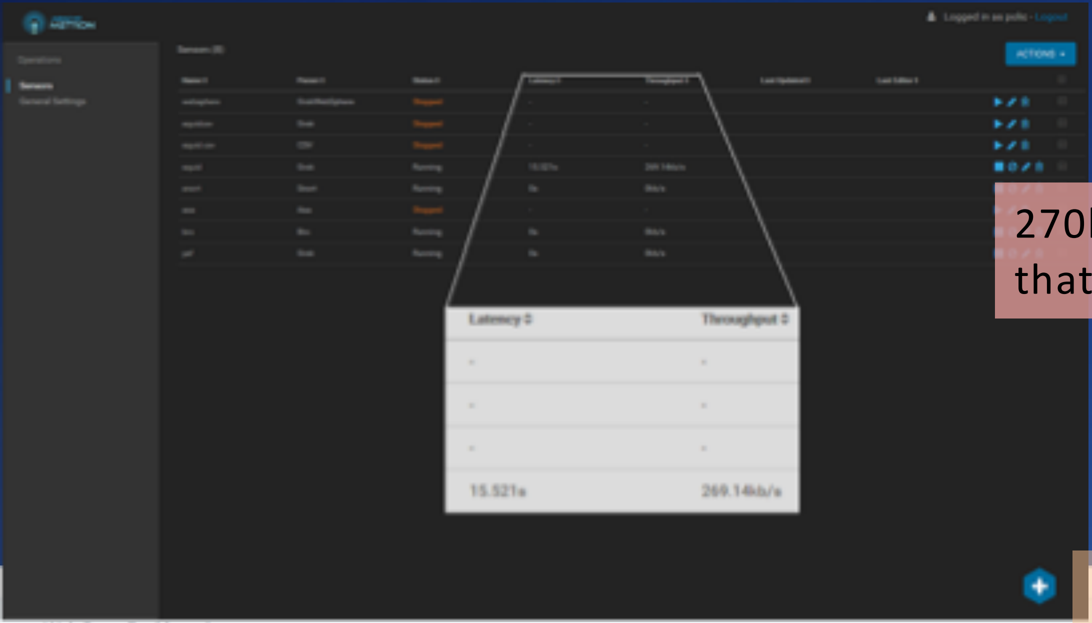
Score	action	id	timestamp	source:type	ip_src_addr	ip_dst_addr	host	alert_status	bytes
-	TCP_TUNNEL	4d5976cb-8...73a3d96c43	2018-03-22 13:05:56	squid	[redacted]	[redacted]		NEW	1775074811
-	TCP_TUNNEL	a37b7382-a...8c97fa5185	2018-03-22 11:07:31	squid	[redacted]	[redacted]		NEW	113140
-	TCP_TUNNEL	9f72e370-a...d0fd73be15	2018-03-22 11:44:57	squid	[redacted]	[redacted]		NEW	6544
-	TCP_TUNNEL	596a245e-5...d787c42971	2018-03-27 11:54:17	squid	[redacted]	[redacted]		NEW	6310
-	TCP_TUNNEL	4795e7b0-0...ff1a0d8f33	2018-03-22 11:44:58	squid	[redacted]	[redacted]		NEW	6107
-	TCP_TUNNEL	becfe7b4-8...10a88f4da7	2018-03-22 12:32:42	squid	[redacted]	[redacted]		NEW	5834
-	TCP_TUNNEL	a29bbd88-c...4fe6264593	2018-03-22 12:39:32	squid	[redacted]	[redacted]		NEW	5834
-	TCP_TUNNEL	c694ae1d-0...d578c9c307	2018-03-22 07:34:05	squid	[redacted]	[redacted]		NEW	5802
-	TCP_TUNNEL	7e1538de-7...cca5fa99b5	2018-03-22 12:39:43	squid	[redacted]	[redacted]		NEW	258

Apache Metron provides a scalable advanced security analytics framework built with the Hadoop Community evolving from the Cisco OpenSOC Project. A cyber security application framework that provides organizations the ability to detect cyber anomalies and enable organizations to rapidly respond to identified anomalies.

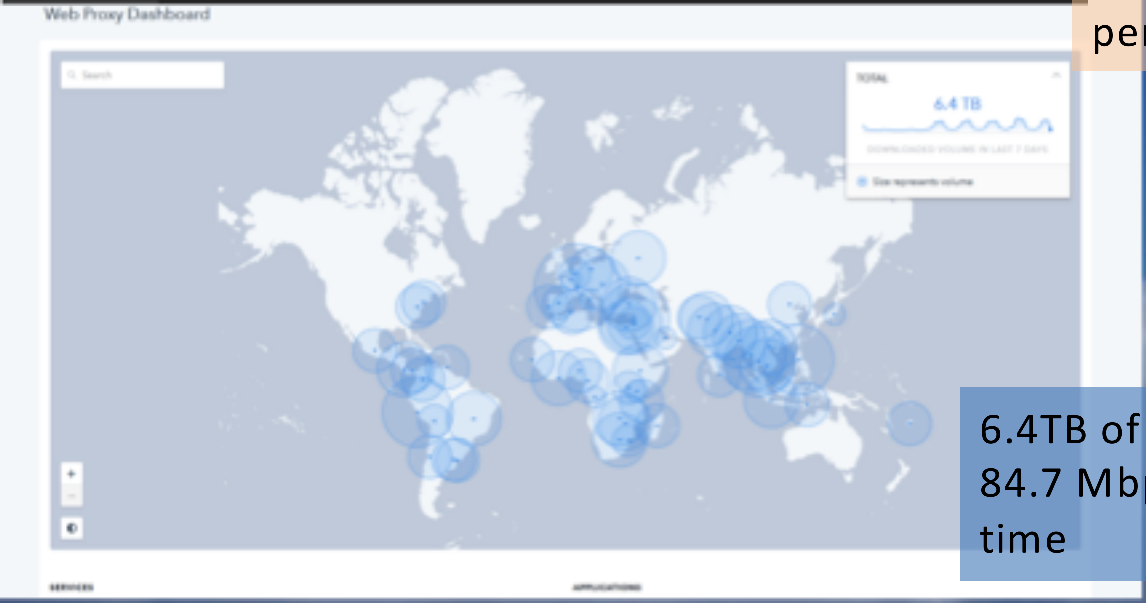
OpenSoC origins: https://blogs.cisco.com/openatcisco/announcing-opensoc-1

OpenSoC community: http://opensoc.github.io/
Apache Metron: http://metron.apache.org/

Unknown unknowns



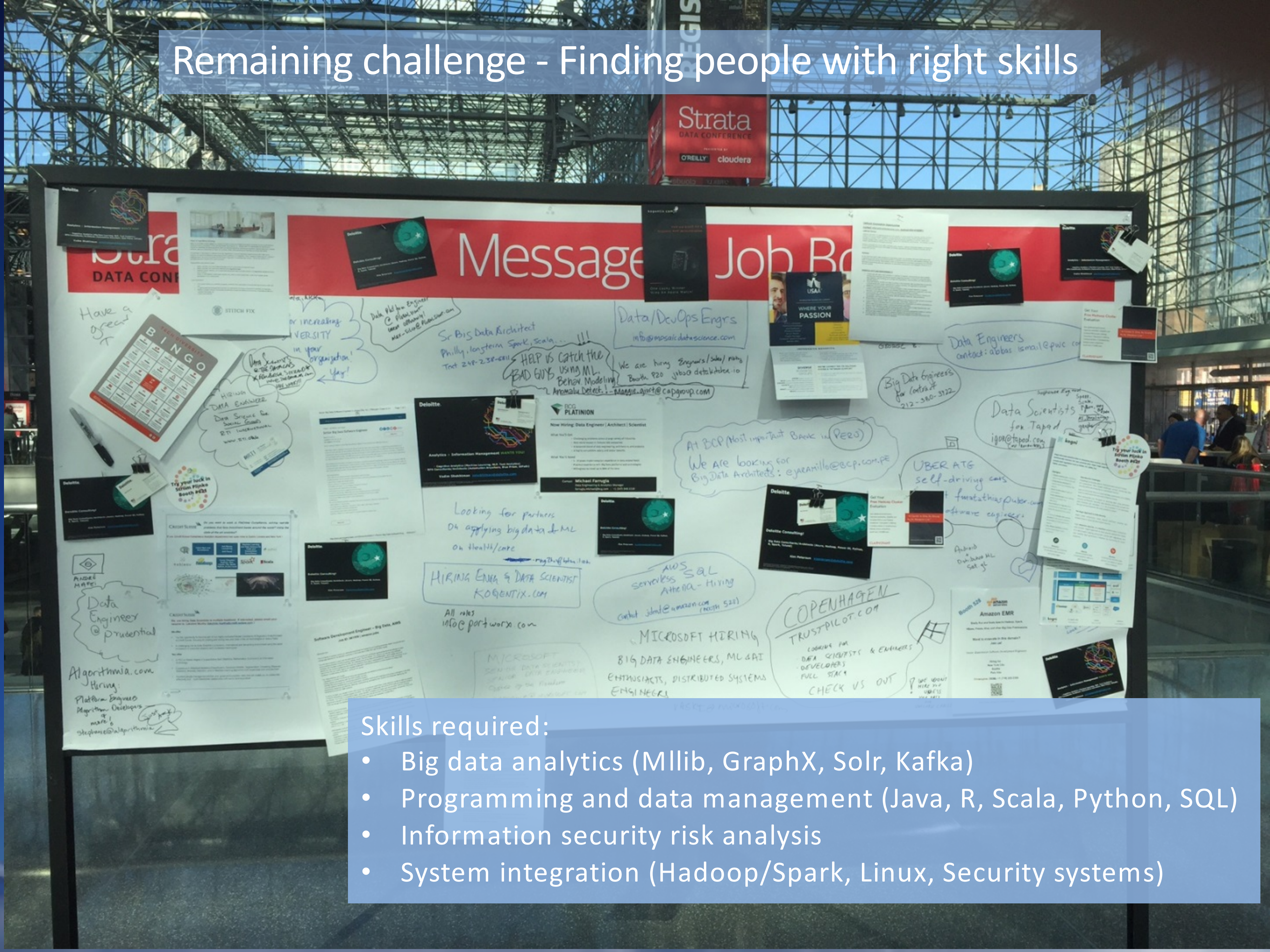
270kbps of unknown unknowns remain that represent 3% of all Internet traffic!



Internet traffic is analyzed in Network perimeter SoC and in End-point SoC

6.4TB of weekly Internet traffic represents 84.7 Mbps of signals to be analyzed in real time

Remaining challenge - Finding people with right skills



Strata
DATA CONFERENCE
O'REILLY cloudera
Message Job Bo
Have a great
BINGO
HIRING DATA ENGINEER
Sr Big Data Architect
HELP US Catch the BAD GUYS using ML, Behav. Modeling, Anomaly Detect.!
Data/DevOps Engrs
info@mosaicdatascience.com
WHERE YOUR PASSION
Data Engineers
Big Data Engineers for Contract
Data Scientists for Tapad
igor@tapad.com
At BCP (Most important Bank in PERU) We Are looking for Big Data Architects
UBER ATG self-driving cars
Try your luck in Scrum Plinko Booth #421
Looking for partners on applying big data & ML on Health/care
HIRING ENGR & DATA SCIENTIST KOGENTIX.COM
All roles info@portworx.com
AWS Serverless SQL Athena - Hiring
MICROSOFT HIRING BIG DATA ENGINEERS, ML & AI ENTHUSIASTS, DISTRIBUTED SYSTEMS ENGINEERS
COPENHAGEN TRUSTPILOT.COM
LOOKING FOR DATA SCIENTISTS & ENGINEERS DEVELOPERS FULL STACK
CHECK US OUT
Data Engineer @ Prudential
Algorithmia.com Hiring Platform Engineers Algorithm Developers
Amazon EMR
BCG PLATINION Now Hiring: Data Engineer | Architect | Scientist
Deloitte.

Skills required:

- Big data analytics (Mllib, GraphX, Solr, Kafka)
- Programming and data management (Java, R, Scala, Python, SQL)
- Information security risk analysis
- System integration (Hadoop/Spark, Linux, Security systems)

Next step - Migrating Security Data Analytics to Cloud Computing Services

- Increasing data processing capacity while maintaining recurrent expenditure levels and reducing capital expenditures (elastic resource provisioning)
- Reducing infrastructure and platform administration overhead (Security data analysts shouldn’t spend time on infrastructure administration)
- Increasing availability
- Implementing data segregation for security purposes (integrity of audit logs)
- Standardization facilitates managed services (more service providers)

No but really, why cloud?

- Lack of internal skills for big data analytics
- Zero-growth budget and increasing cybersecurity related work (95% vs. 5% threat detection)
- Externalization (outsourcing) of security data management – decentralized security operations
- Risk shifting to end users and end-point devices (they are mobile and globally dispersed)
- “Cloud first” organizational IT strategy – “crown jewels” data will be in the cloud
- Need to have 24/7 near real time alerting for unknown unknowns (unforeseen and unexpected events) in order to have a proactive security risk management

Migration requirements

- Protect investment in in-house developed solution (Data LTE procedures, ML code, data visualization)
- Open architecture (data exchange, taxonomies, interoperability with 3rd party SaaS providers)
- Dynamic resource provisioning
- Availability of managed services for IaaS, PaaS
- Security requirements (data segregation, encryption, access auditing, log monitoring, high-availability)
- Specific organizational legal and contractual requirements (flexibility to select data jurisdictions/regions)

Proof of Concept (PoC) methodology

- Stage 1 PoC – feasibility study
  - Replicate current on-premise platform into the cloud with light-weight optimization
  - Migrate test data sets for batch processing
  - Run 2 processing ML tests (LoF and k-means clustering) and compare metrics with on-premise results
  - Test data stream indexing, classification and search
  - Iterate through all selected SaaS vendors
  - Measure resources and costs
- Stage 2 PoC – resource scaling and optimization
  - Heavier optimization of all data management phases
  - Larger and more diverse data sets
  - More data analytics ML tests
  - Model resources and costs
  - Assess managed services (availability, quality and costs)

Feasibility study findings

- Not all regions (data centers) allow dynamic resource provisioning and high-availability (important for data streaming)
- Optimization expertise consultancy is expensive and scarce
- Integration with 3rd party SaaS providers varies significantly which can lead to 4x higher dynamic resource utilization and costs!
- 3rd party SaaS providers don’t support all PaaS data structures
- Multiple SaaS providers hybrid solution might be the best option
- Complex formula for Total Cost of Ownership (TCO) calculations:

data storage (including backup/restores and archiving) + data processing + bandwidth + operations + optimization + development - risk (including lost opportunities)

Contacts and further information

Dr. Viktor Polic

vpolic@cybersymbiosis.com

https://ch.linkedin.com/in/viktor-polic-891a1a145

https://twitter.com/ViktorPolic