RSAConference2015

San Francisco | April 20-24 | Moscone Center

CHANGE

Challenge today's security thinking

SESSION ID: STR-W03

Data Science Transforming Security Operations

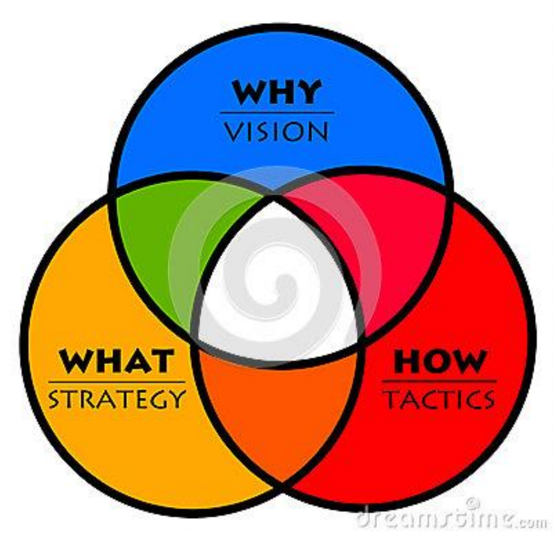
Alon Kaufman Ph.D.

Director Data Science & Innovation
RSA

Agenda

Transforming Security Operations with Data Science

- The **Vision**: Where we should head & Why
- The **Strategy**: What we need to achieve
- The **Tactics**: How we get there



Key message: Data science is a key methodology and technology in this transformation, its not merely a feature….

Security Operation – Current Operation Model

Threat Landscape

More aggressive

More complex

From everywhere

On everything

Current Security Operations

Collect

Detect

Investigate

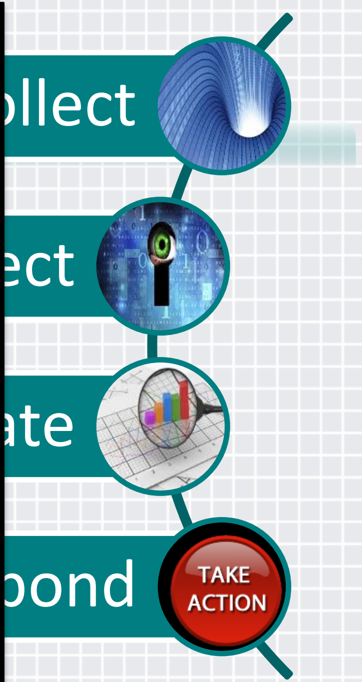
Respond

TAKE
ACTION

Security Operation – Current Operation Model

Threat Landscape

Current Security Operations



Current Approach in Failing

2015

Information security incidents:

42.8 M

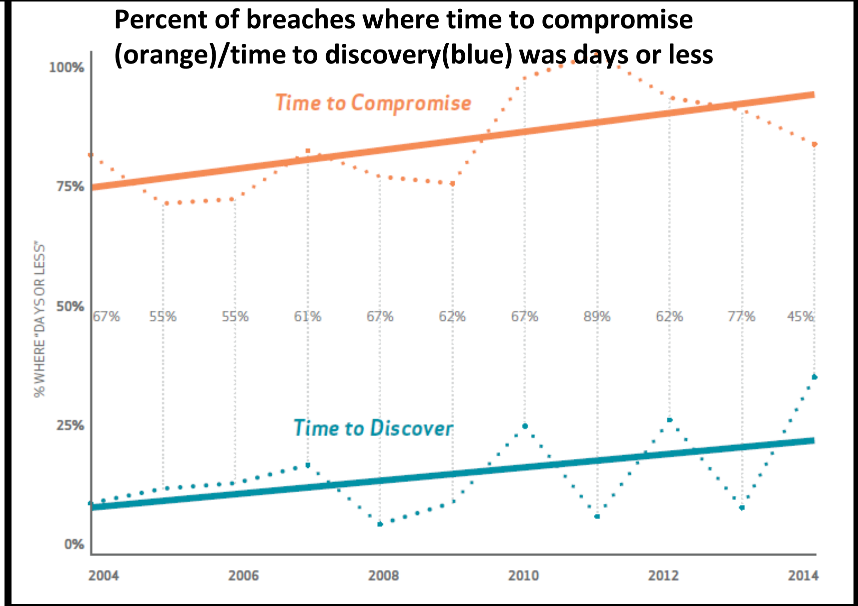
↑48%

Confirmed breaches:

2122

↑55%

Percent of breaches where time to compromise (orange)/time to discovery(blue) was days or less



Failures of the security industry, Yoran, Robertson, 2015

Verizon, 2005 DBIR

We are Being Defeated, What's Happening?

- We have the technology, the brains, the funding….

The attackers operate in a known environment!

The attackers job is more predictable!!

Their lives are easier!!!

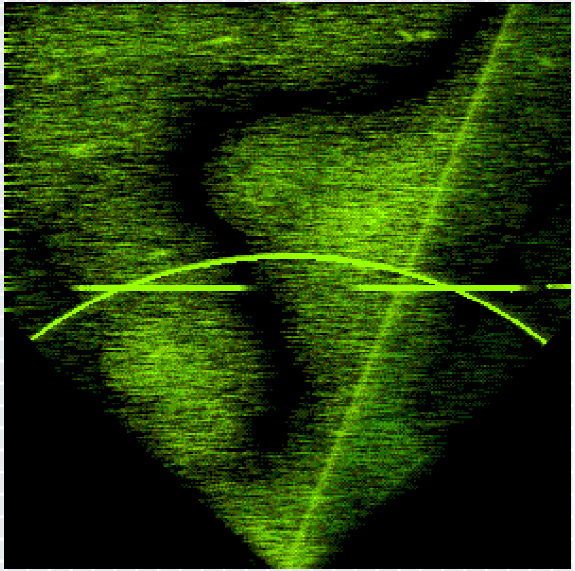
Our operations are slow… reactive…

We are running into a skill gap/fatigue…

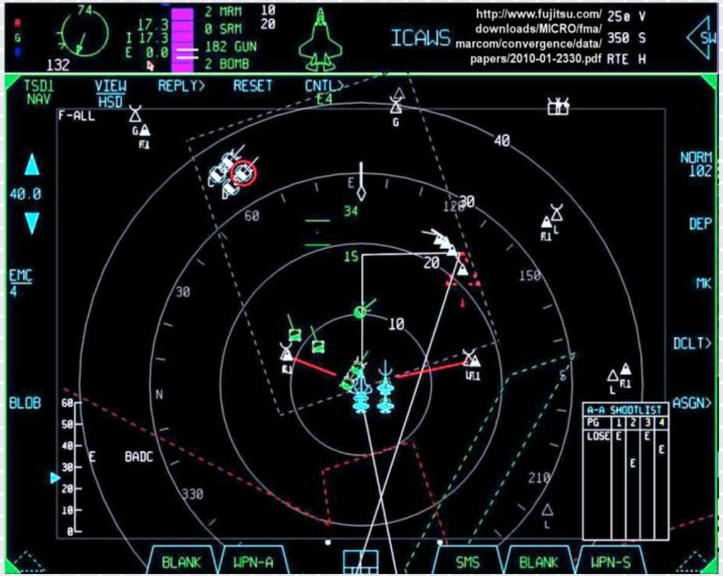
We don't make their life hard enough…

Technology at the Aid of Human Operations

1965 Aviation Radar



2015 Tactical Situation Display



http://www.fujitsu.com/downloads/MICRO/fma/marcom/convergence/data/papers/2010-01-2330.pdf

Transforming the Role of the Human Operations Behind the Machine

What's the "driver's" role?

What's the "Pilot's" role?

The Human role:

Leaving to the human the intentions & final decisions!!

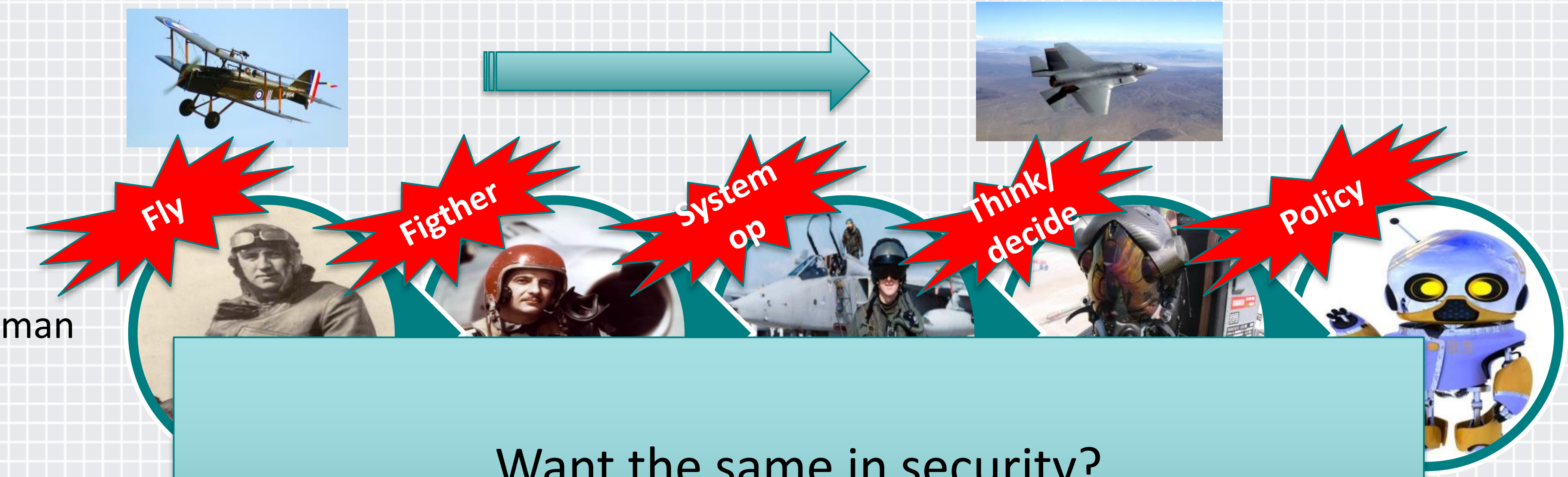
What Happened to The Machine Operator

Human

Platform	“Stick & throttle” Gun	Jet engines Missiles	Radar BVR missiles RWR	Detect threat Identify danger ECM Select weapon Human decision	…

*Beyond View Range, **Radar Warning Receiver ,***Electronic CounterMeasure

What Happened to The Machine Operator



Fly

Figther

System op

Think/ decide

Policy

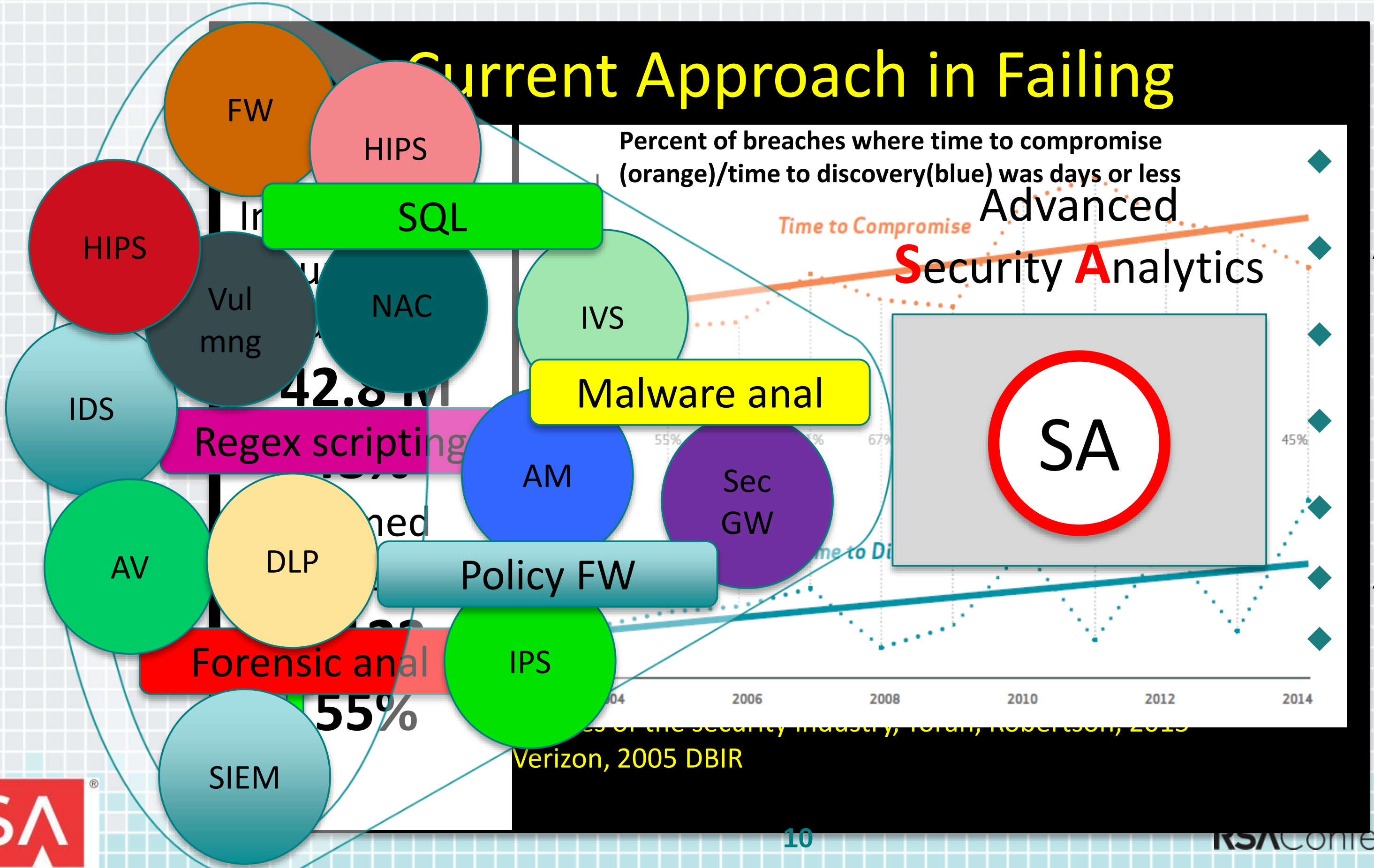
Human

Platform

Want the same in security?

Leave thinking and decisions to humans and let the machines and platforms do the rest

It’s Time for the Next Era



Current Approach in Failing

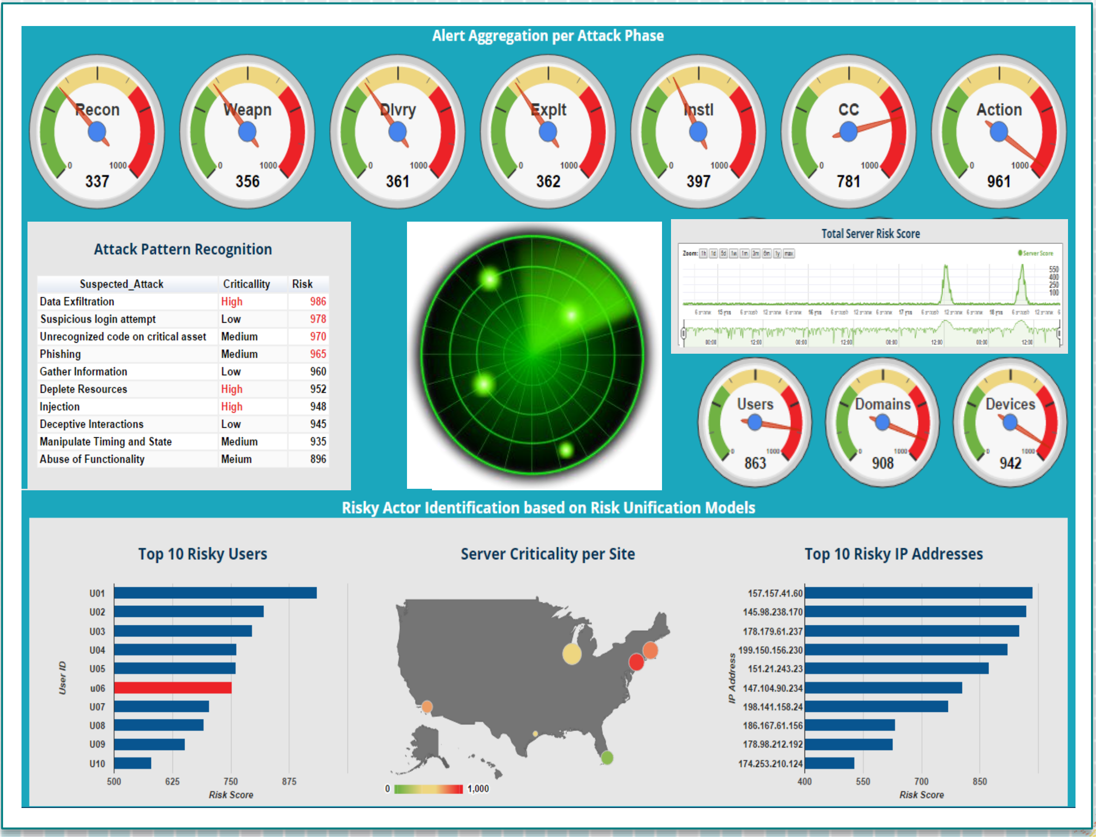
Percent of breaches where time to compromise (orange)/time to discovery(blue) was days or less

Verizon, 2005 DBIR

- Comprehensive
- Accurate
- Proactive
- Faster
- Context aware
- Adaptive
- Leverage intelligence

The Next Era: Advanced Security Analytics

- **Comprehensive**
  - All possible inputs
  - Going beyond the known
- **Proactive / Preventive**
  - Automatic
  - Actionable insights & recommendation
- **Human intelligence**
  - Knowledge
  - Sharing
  - Intentions & decisions

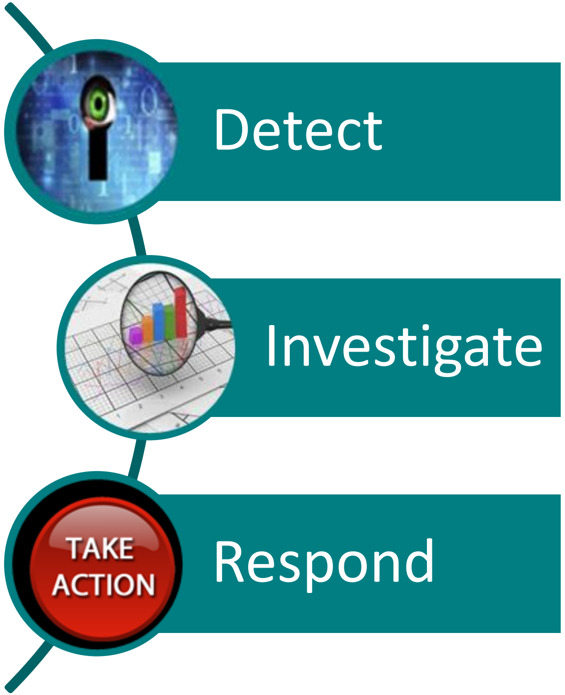


Attack Pattern Recognition

Suspected_Attack	Criticality	Risk
Data Exfiltration	High	986
Suspicious login attempt	Low	978
Unrecognized code on critical asset	Medium	970
Phishing	Medium	965
Gather Information	Low	960
Deplete Resources	High	952
Injection	High	948
Deceptive Interactions	Low	945
Manipulate Timing and State	Medium	935
Abuse of Functionality	Meium	896

The Next Era: Advanced Security Analytics

- **Comprehensive**
  - All possible inputs
  - Going beyond the known
- **Proactive / Preventive**
  - Automatic
  - Actionable insights & recommendation
- **Human intelligence**
  - Knowledge
  - Sharing
  - Intentions & decisions



Detect

- Attacks vs Alerts
- Prioritized
- Immediate

Investigate

- Hierarchical model
- All prepared
- Risk & impact

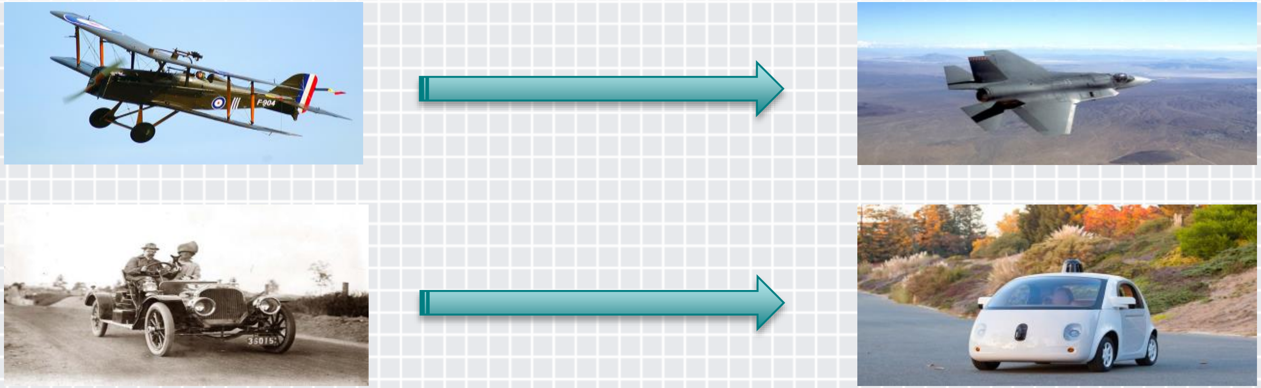
Respond

- Automatic
- Self learning
- Recommend

We Need to Transform…

So ***what*** is needed to get there?

We Need to Transform…



So *what* is needed to get there?

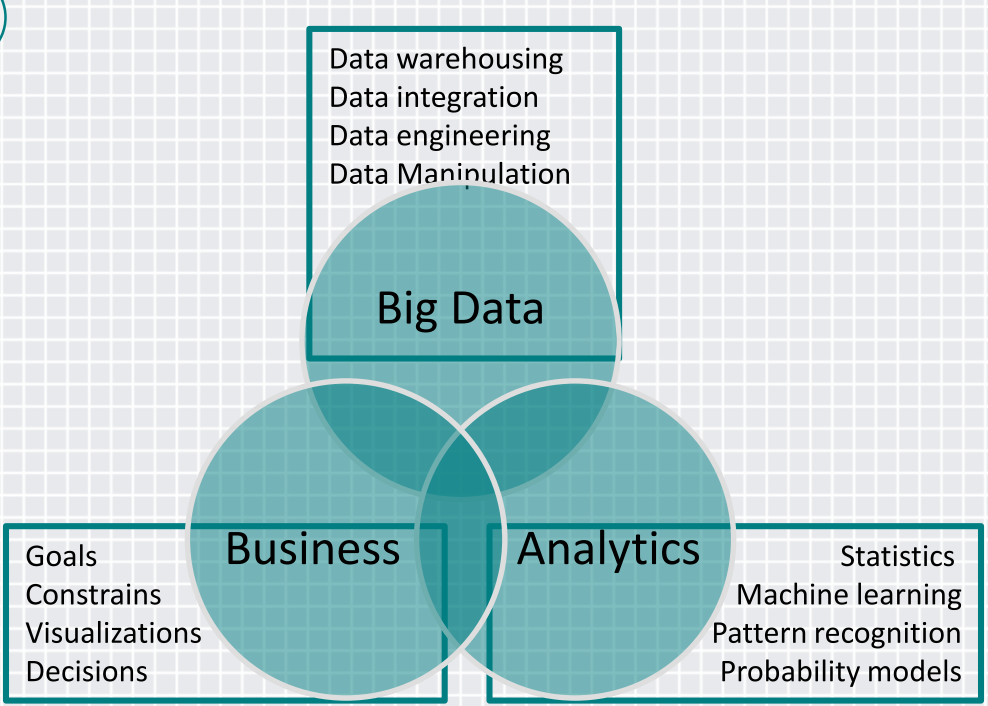
We have the **DATA** – We know the **PROBLEM**…

we need to reveal the insights & provide actionable outcomes

Data Science!!!

What is Data Science in 47 Seconds

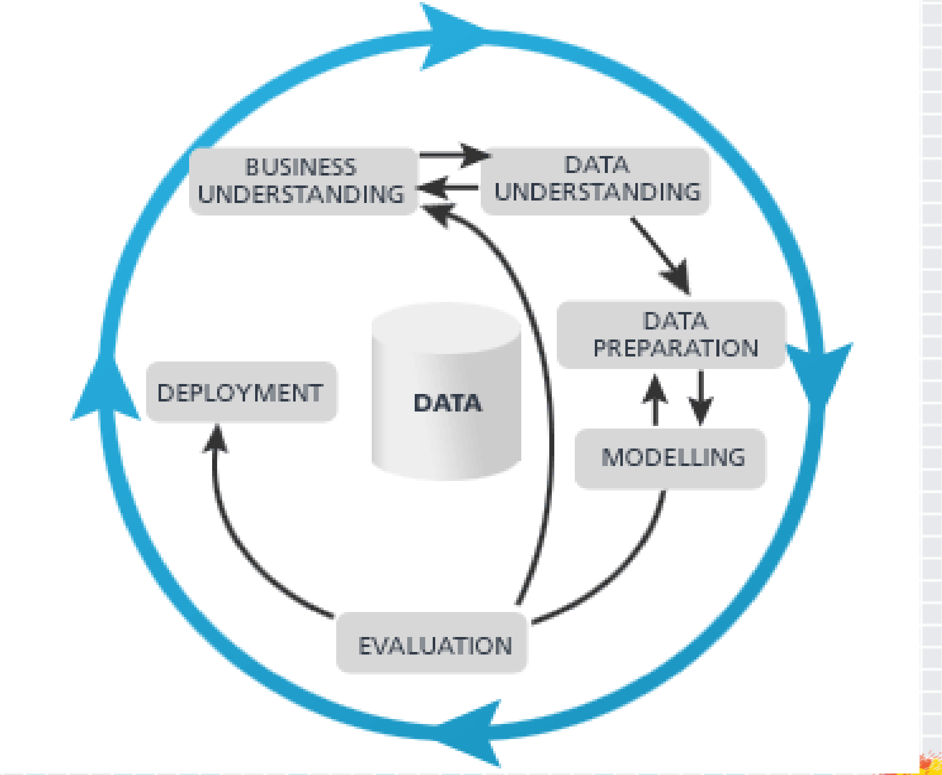
*The use of techniques as **statistics** and **machine learning** on big multi-structured data, to identify **correlations** and causal **relationships**, **classify** and **predict** events, identify **patterns** and **anomalies**, **infer** probabilities and interest, with the goal of extracting meaning from data and creating valuable products.*



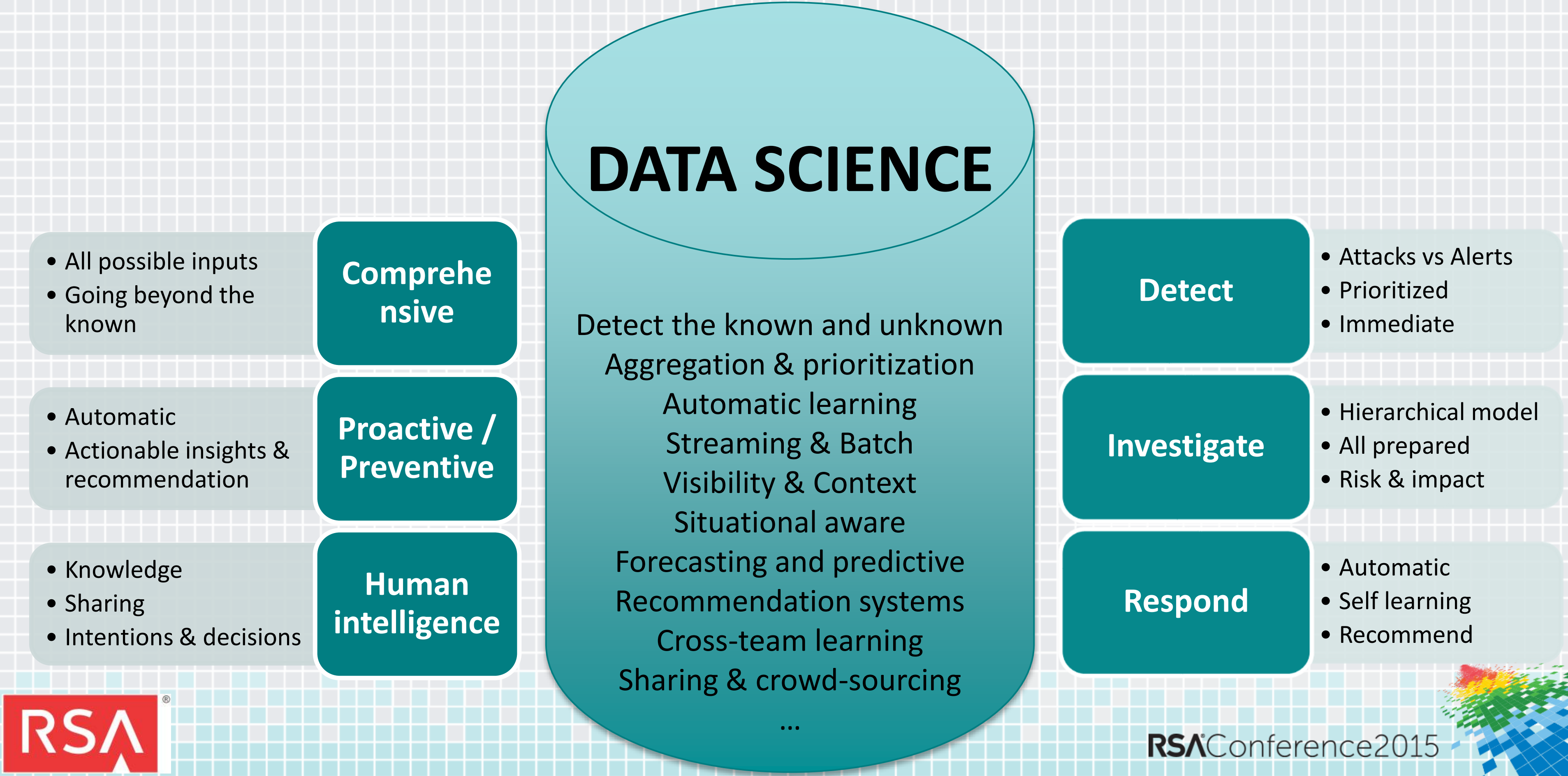
The Art of solving business problems utilizing the available data

Data Science Applicative Practice

- Data Science is no magic!
  - Understand your problem
  - Learn your data
  - Create valuable indicators
  - Apply the right modeling techniques!
  - Measure & Success
- Joint iterative work with domain experts and SMEs



The Data is all There -> Reveal the Essence



DATA SCIENCE

Detect the known and unknown
Aggregation & prioritization
Automatic learning
Streaming & Batch
Visibility & Context
Situational aware
Forecasting and predictive
Recommendation systems
Cross-team learning
Sharing & crowd-sourcing
…

Comprehensive

- All possible inputs
- Going beyond the known

Proactive / Preventive

- Automatic
- Actionable insights & recommendation

Human intelligence

- Knowledge
- Sharing
- Intentions & decisions

Detect

- Attacks vs Alerts
- Prioritized
- Immediate

Investigate

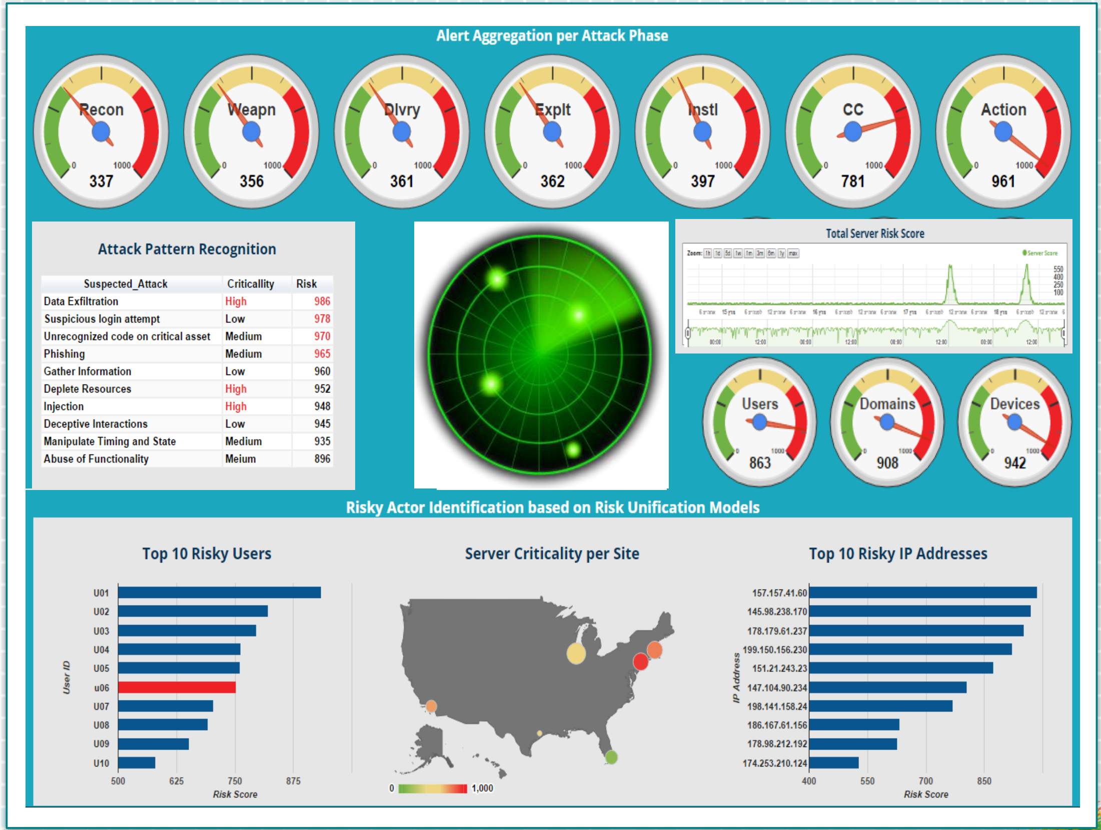
- Hierarchical model
- All prepared
- Risk & impact

Respond

- Automatic
- Self learning
- Recommend

Examples of Data Science in the Transformation

- **Detect**
  - Attacks vs Alerts
  - Prioritized
  - Immediate
- **Investigation**
  - Hierarchical model
  - All prepared
  - Risk & impact
- **Respond**
  - Automatic
  - Self learning
  - Recommend



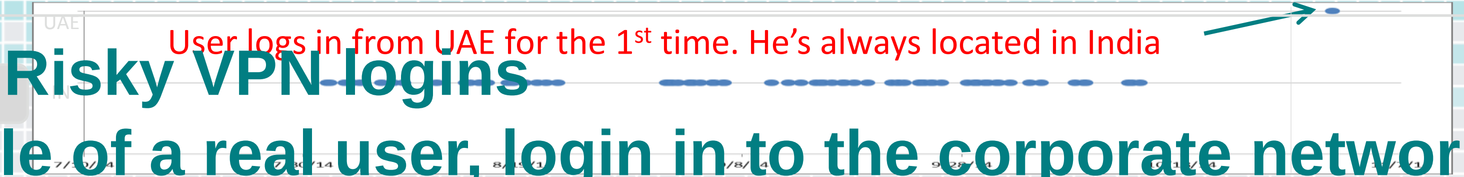
Attack Pattern Recognition

Suspected_Attack	Criticallity	Risk
Data Exfiltration	High	986
Suspicious login attempt	Low	978
Unrecognized code on critical asset	Medium	970
Phishing	Medium	965
Gather Information	Low	960
Deplete Resources	High	952
Injection	High	948
Deceptive Interactions	Low	945
Manipulate Timing and State	Medium	935
Abuse of Functionality	Meium	896

Detect Risky VPN logins

Example of a real user, login in to the corporate network

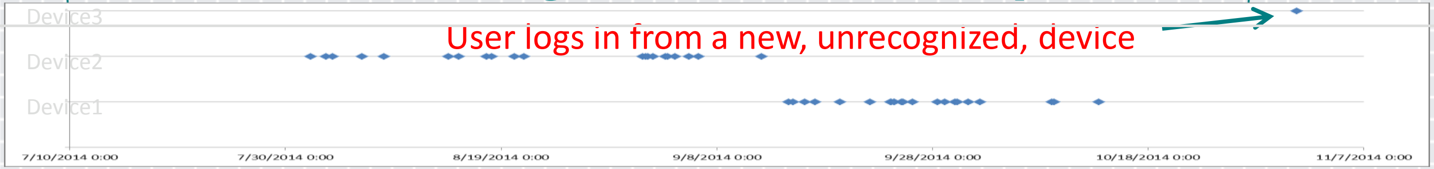
Country



User logs in from UAE for the 1st time. He's always located in India

Score: **92**

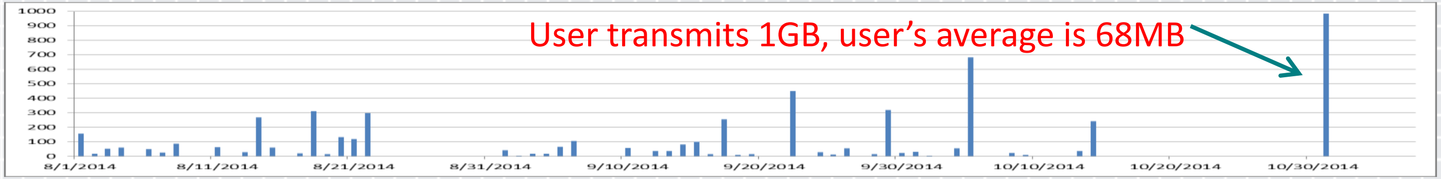
Device



User logs in from a new, unrecognized, device

Score: **90**

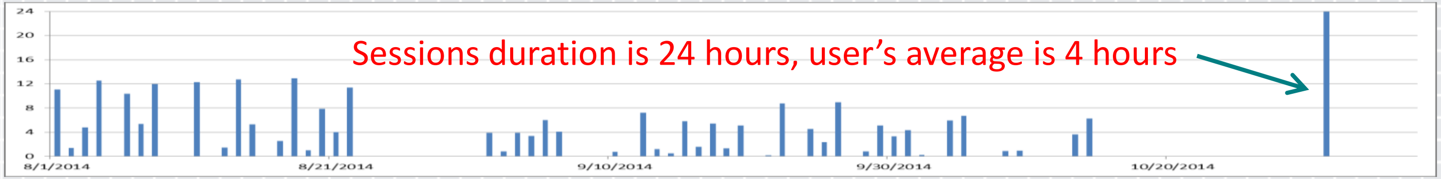
Transmitted Data [MB]



User transmits 1GB, user's average is 68MB

Score: **93**

Session Duration [hr]

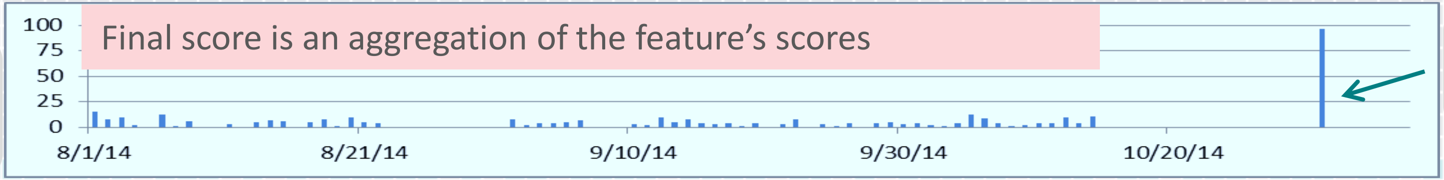


Sessions duration is 24 hours, user's average is 4 hours

Score: **84**

Many more

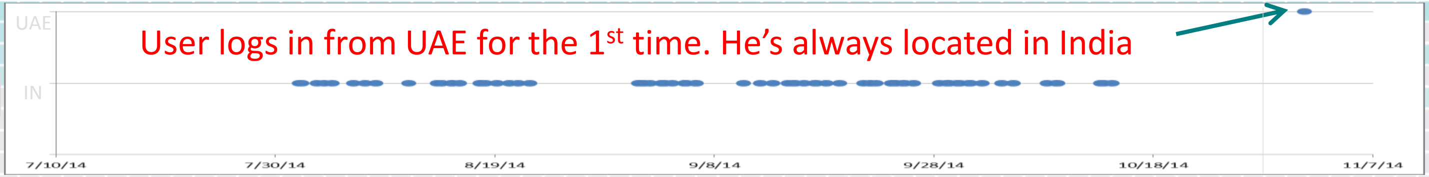
Final Score



Final score is an aggregation of the feature's scores

Aggregate Score: **98**

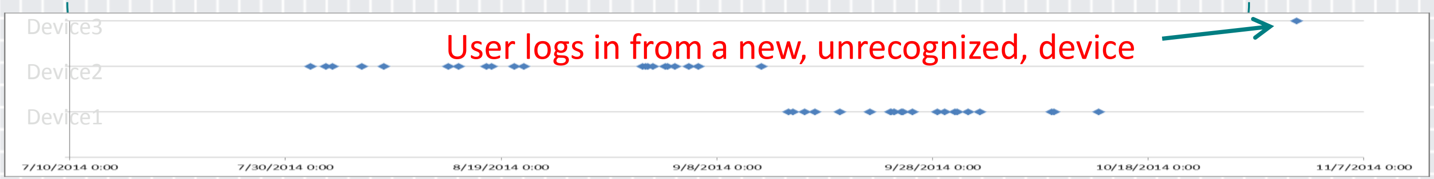
Country



User logs in from UAE for the 1st time. He's always located in India

Score: **92**

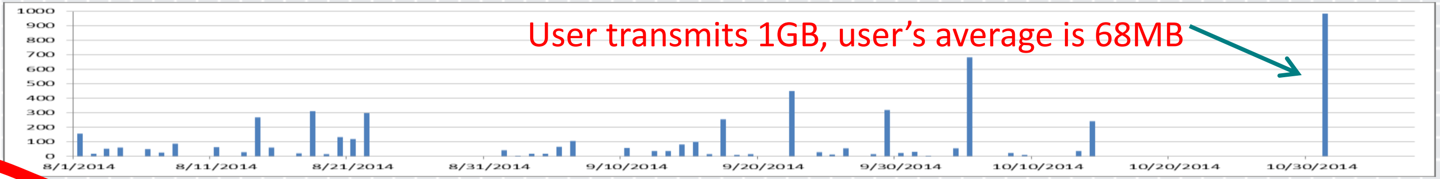
Device



User logs in from a new, unrecognized, device

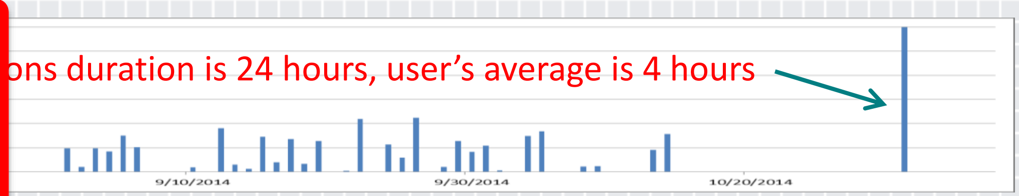
Score: **90**

Transmitted Data [MB]



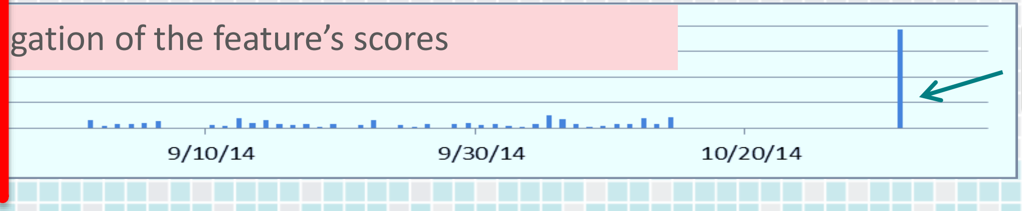
User transmits 1GB, user's average is 68MB

Score: **93**



ons duration is 24 hours, user's average is 4 hours

Score: **84**



gation of the feature's scores

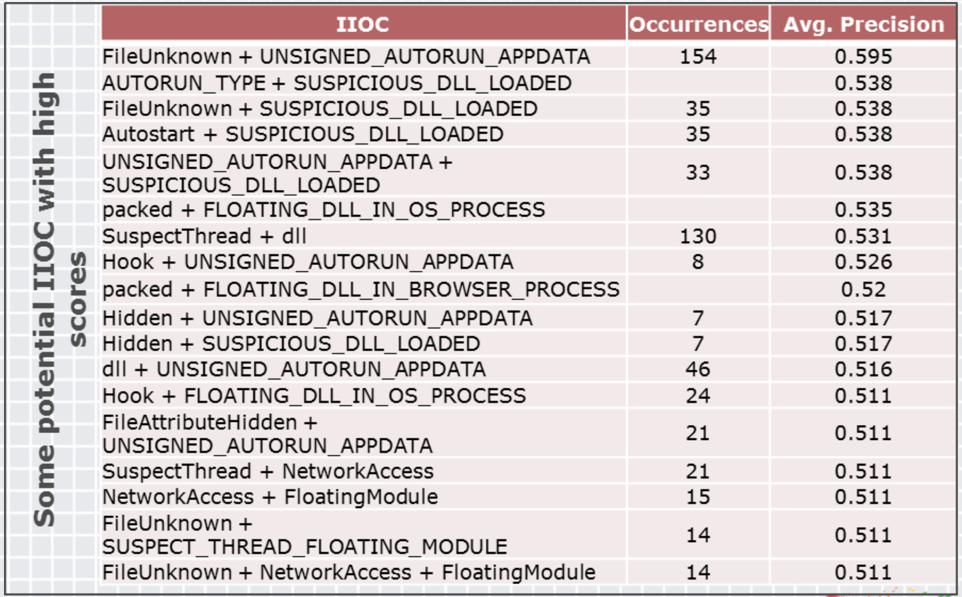
Aggregate Score: **98**

- Based on nested anomaly detection models
- User/Group behavior analysis
- Ranking based on statistical significance

Recommending New IOCs

- When the analyst creates a new IOC the system can recommend:
  - How strong is this new IOC by itself?
  - Is it similar to already existing IOC?
  - What is the best threshold setting (if applicable)?
  - Recommends new IoC

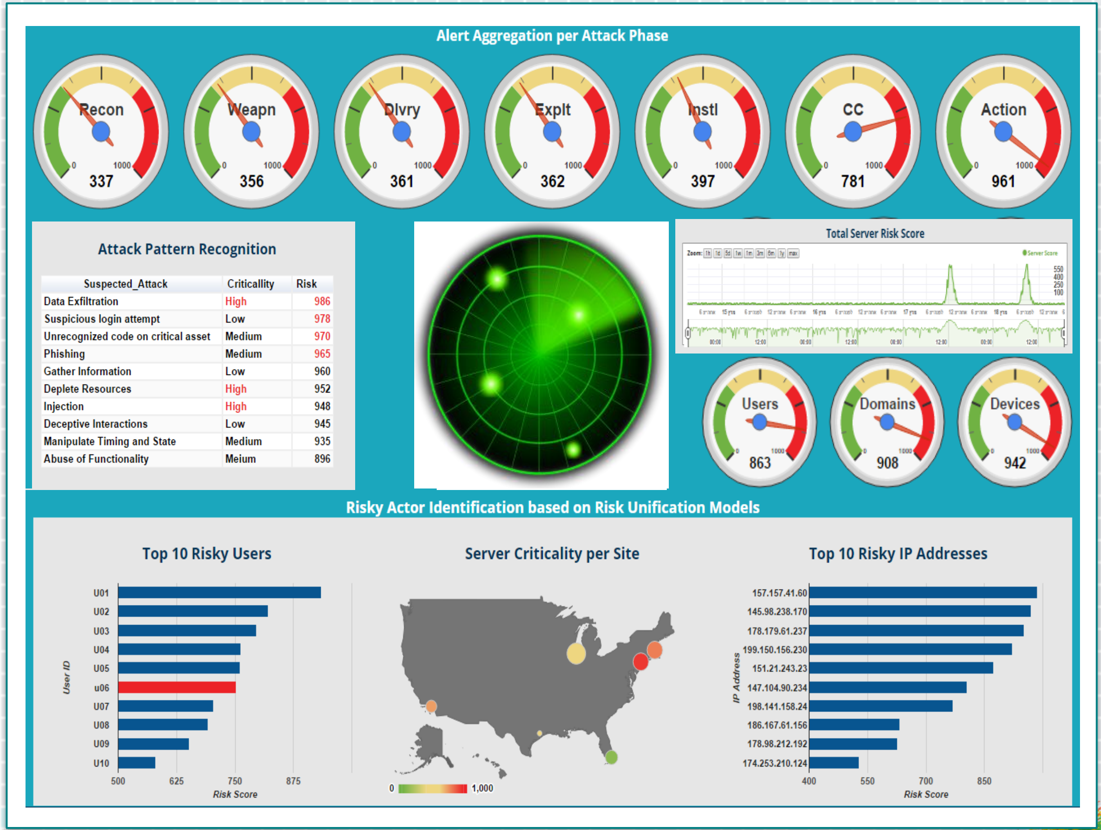
New Recommended IoCs, based on combining existing IoCs



	IIOC	Occurrences	Avg. Precision
Some potential IIOC with high scores	FileUnknown + UNSIGNED_AUTORUN_APPDATA	154	0.595
	AUTORUN_TYPE + SUSPICIOUS_DLL_LOADED		0.538
	FileUnknown + SUSPICIOUS_DLL_LOADED	35	0.538
	Autostart + SUSPICIOUS_DLL_LOADED	35	0.538
	UNSIGNED_AUTORUN_APPDATA + SUSPICIOUS_DLL_LOADED	33	0.538
	packed + FLOATING_DLL_IN_OS_PROCESS		0.535
	SuspectThread + dll	130	0.531
	Hook + UNSIGNED_AUTORUN_APPDATA	8	0.526
	packed + FLOATING_DLL_IN_BROWSER_PROCESS		0.52
	Hidden + UNSIGNED_AUTORUN_APPDATA	7	0.517
	Hidden + SUSPICIOUS_DLL_LOADED	7	0.517
	dll + UNSIGNED_AUTORUN_APPDATA	46	0.516
	Hook + FLOATING_DLL_IN_OS_PROCESS	24	0.511
	FileAttributeHidden + UNSIGNED_AUTORUN_APPDATA	21	0.511
	SuspectThread + NetworkAccess	21	0.511
	NetworkAccess + FloatingModule	15	0.511
	FileUnknown + SUSPECT_THREAD_FLOATING_MODULE	14	0.511
	FileUnknown + NetworkAccess + FloatingModule	14	0.511

Examples of Data Science in the Transformation

- **Detect**
  - Attacks vs Alerts
  - Prioritized
  - Immediate
- **Investigation**
  - Hierarchical model
  - All prepared
  - Risk & impact
- **Respond**
  - Automatic
  - Self learning
  - Recommend



Alert Aggregation per Attack Phase

Attack Pattern Recognition

Suspected_Attack	Criticallity	Risk
Data Exfiltration	High	986
Suspicious login attempt	Low	978
Unrecognized code on critical asset	Medium	970
Phishing	Medium	965
Gather Information	Low	960
Deplete Resources	High	952
Injection	High	948
Deceptive Interactions	Low	945
Manipulate Timing and State	Medium	935
Abuse of Functionality	Meium	896

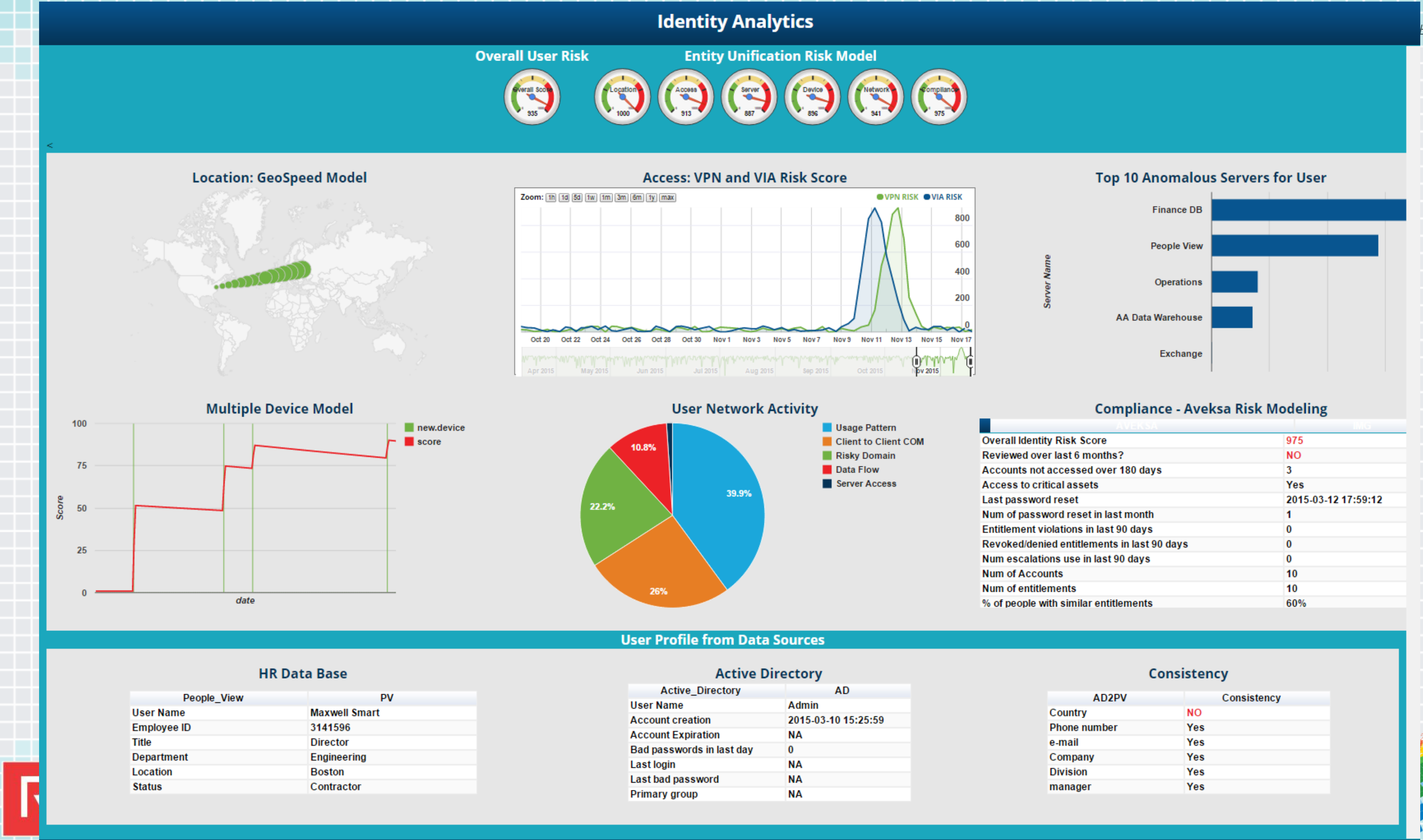
Total Server Risk Score

Risky Actor Identification based on Risk Unification Models

Top 10 Risky Users

Server Criticality per Site

Top 10 Risky IP Addresses



Identity Analytics

Overall User Risk

Entity Unification Risk Model

Compliance - Aveksa Risk Modeling

Overall Identity Risk Score	975
Reviewed over last 6 months?	NO
Accounts not accessed over 180 days	3
Access to critical assets	Yes
Last password reset	2015-03-12 17:59:12
Num of password reset in last month	1
Entitlement violations in last 90 days	0
Revoked/denied entitlements in last 90 days	0
Num escalations use in last 90 days	0
Num of Accounts	10
Num of entitlements	10
% of people with similar entitlements	60%

User Profile from Data Sources

HR Data Base

People_View	PV
User Name	Maxwell Smart
Employee ID	3141596
Title	Director
Department	Engineering
Location	Boston
Status	Contractor

Active Directory

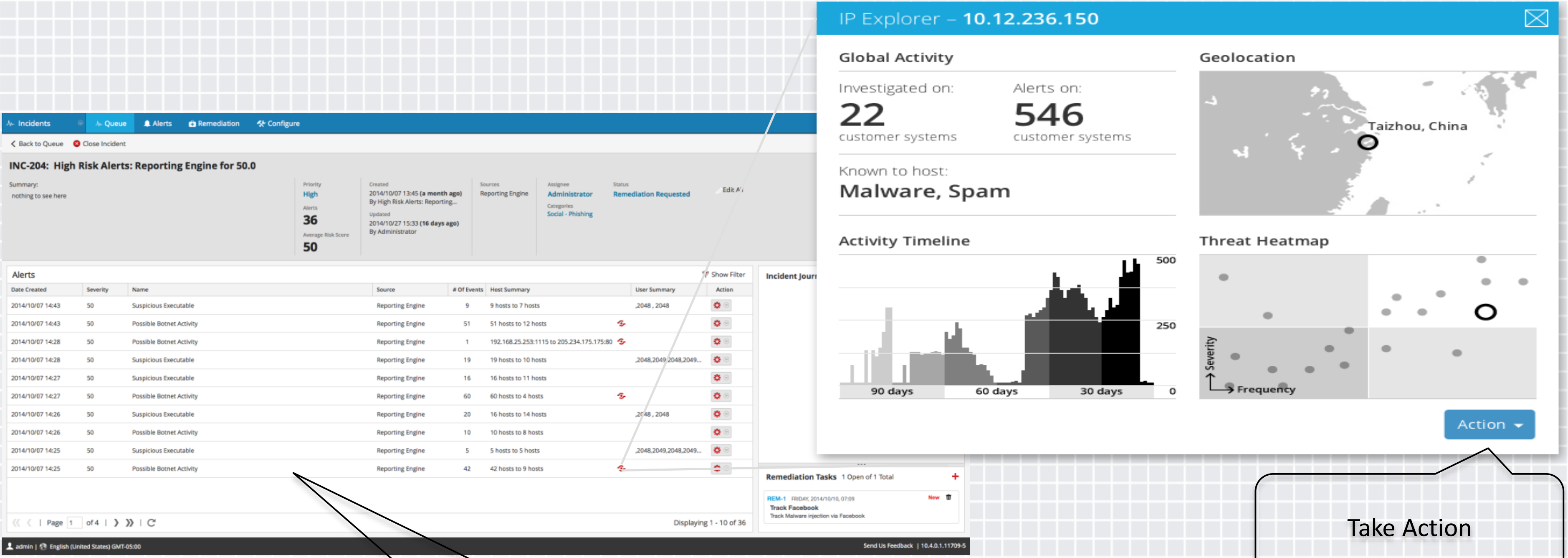
Active_Directory	AD
User Name	Admin
Account creation	2015-03-10 15:25:59
Account Expiration	NA
Bad passwords in last day	0
Last login	NA
Last bad password	NA
Primary group	NA

Consistency

AD2PV	Consistency
Country	NO
Phone number	Yes
e-mail	Yes
Company	Yes
Division	Yes
manager	Yes

Community Based Data Enrichment

- Let the analyst know what the community thinks about the IP



INC-204: High Risk Alerts: Reporting Engine for 50.0

Summary:
nothing to see here

Priority: High
Alerts: 36
Average Risk Score: 50

Created: 2014/10/07 13:45 **(a month ago)**
By High Risk Alerts: Reporting...

Updated: 2014/10/27 15:33 **(16 days ago)**
By Administrator

Sources: Reporting Engine

Assignee: Administrator
Categories: Social - Phishing

Status: Remediation Requested

Alerts

Date Created	Severity	Name	Source	# Of Events	Host Summary	User Summary	Action
2014/10/07 14:43	50	Suspicious Executable	Reporting Engine	9	9 hosts to 7 hosts	,2048 , 2048	
2014/10/07 14:43	50	Possible Botnet Activity	Reporting Engine	51	51 hosts to 12 hosts		
2014/10/07 14:28	50	Possible Botnet Activity	Reporting Engine	1	192.168.25.253:1115 to 205.234.175.175:80		
2014/10/07 14:28	50	Suspicious Executable	Reporting Engine	19	19 hosts to 10 hosts	,2048,2049,2048,2049...	
2014/10/07 14:27	50	Suspicious Executable	Reporting Engine	16	16 hosts to 11 hosts		
2014/10/07 14:27	50	Possible Botnet Activity	Reporting Engine	60	60 hosts to 4 hosts		
2014/10/07 14:26	50	Suspicious Executable	Reporting Engine	20	16 hosts to 14 hosts	,2048 , 2048	
2014/10/07 14:26	50	Possible Botnet Activity	Reporting Engine	10	10 hosts to 8 hosts		
2014/10/07 14:25	50	Suspicious Executable	Reporting Engine	5	5 hosts to 5 hosts	,2048,2049,2048,2049...	
2014/10/07 14:25	50	Possible Botnet Activity	Reporting Engine	42	42 hosts to 9 hosts		

Displaying 1 - 10 of 36

Remediation Tasks 1 Open of 1 Total

REM-1 FRIDAY, 2014/10/10, 07:09 New
Track Facebook
Track Malware injection via Facebook

admin | English (United States) GMT-05:00

Send Us Feedback | 10.4.0.1.11709-5

IP Explorer – **10.12.236.150**

Global Activity

Investigated on:
22
customer systems

Alerts on:
546
customer systems

Known to host:
Malware, Spam

Geolocation

Activity Timeline

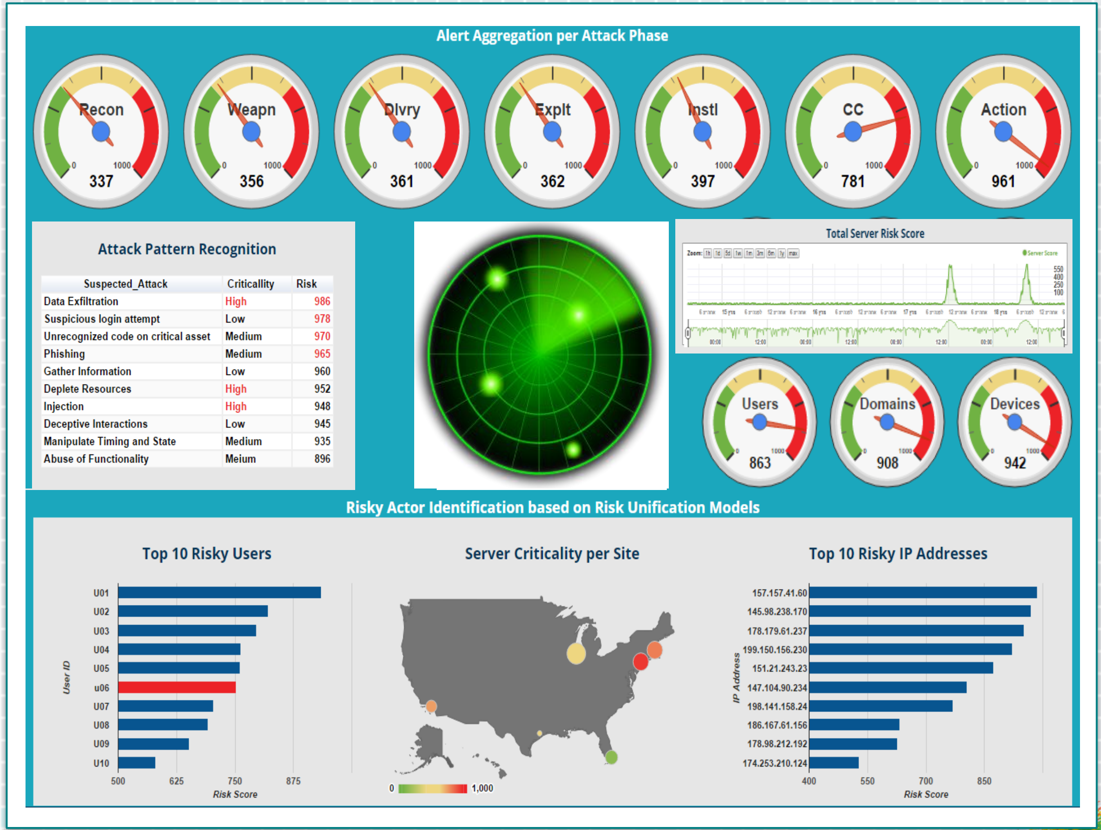
Threat Heatmap

Investigate & Incidents

Take Action

Examples of Data Science in the Transformation

- **Detect**
  - Attacks vs Alerts
  - Prioritized
  - Immediate
- **Investigation**
  - Hierarchical model
  - All prepared
  - Risk & impact
- **Respond**
  - Automatic
  - Self learning
  - Recommend



Alert Aggregation per Attack Phase

Attack Pattern Recognition

Suspected_Attack	Criticallity	Risk
Data Exfiltration	High	986
Suspicious login attempt	Low	978
Unrecognized code on critical asset	Medium	970
Phishing	Medium	965
Gather Information	Low	960
Deplete Resources	High	952
Injection	High	948
Deceptive Interactions	Low	945
Manipulate Timing and State	Medium	935
Abuse of Functionality	Meium	896

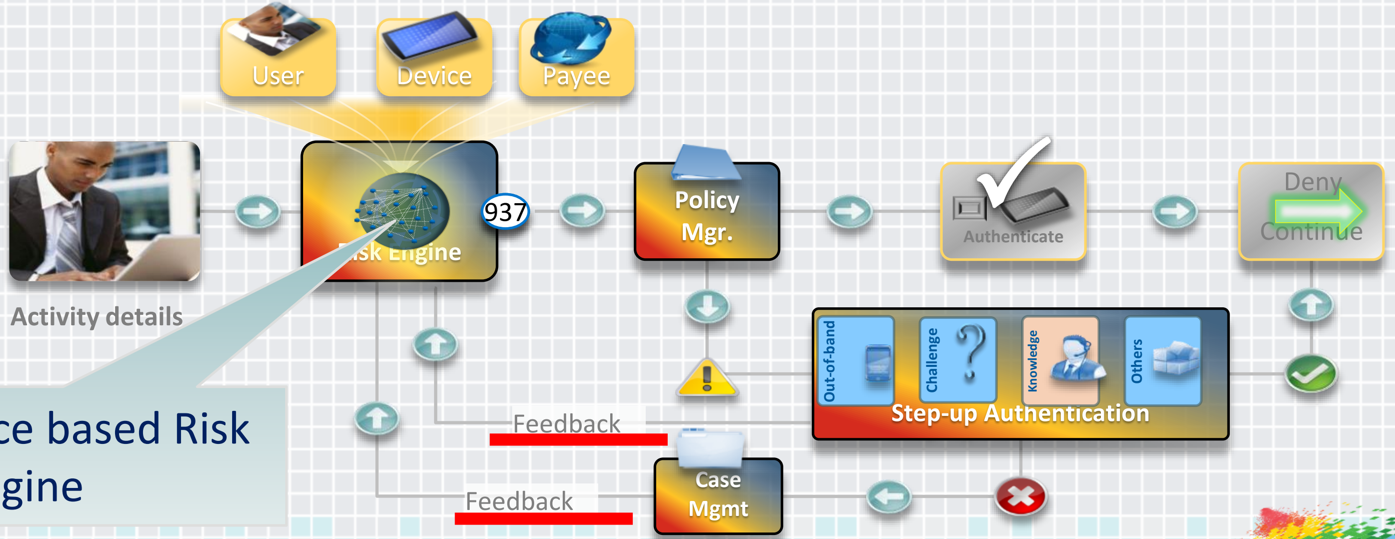
Total Server Risk Score

Risky Actor Identification based on Risk Unification Models

Leaning and Self-Improving Detection

Risk-Based Security and Self-Protection

- Ongoing, automatic self-learning fraud detection model
- Real-time decisions/actions based on risk scores



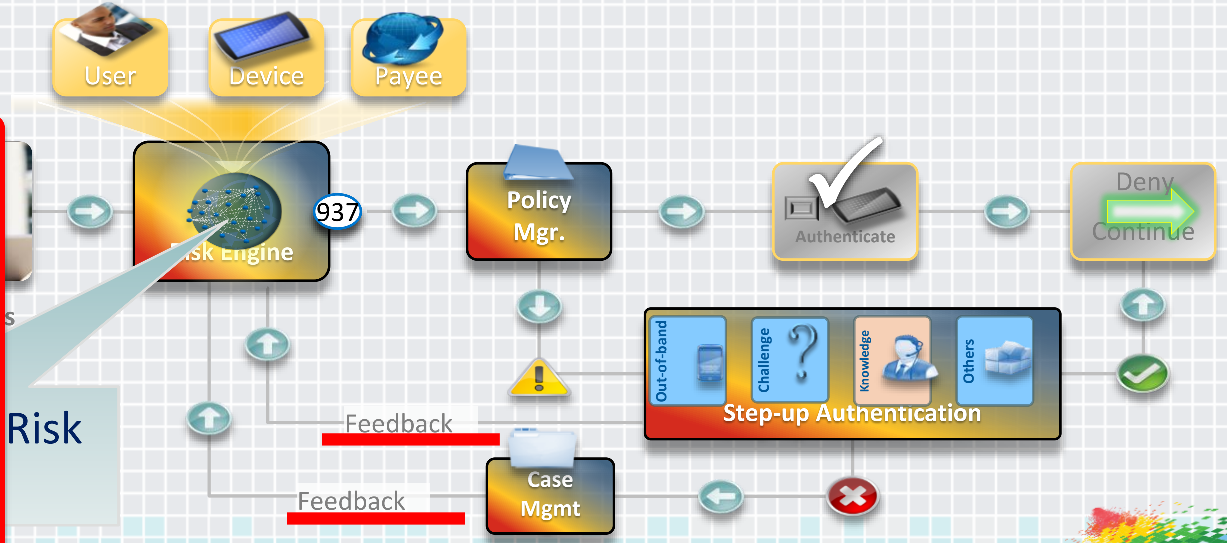
Data Science based Risk Engine

Leaning and Self-Improving Detection

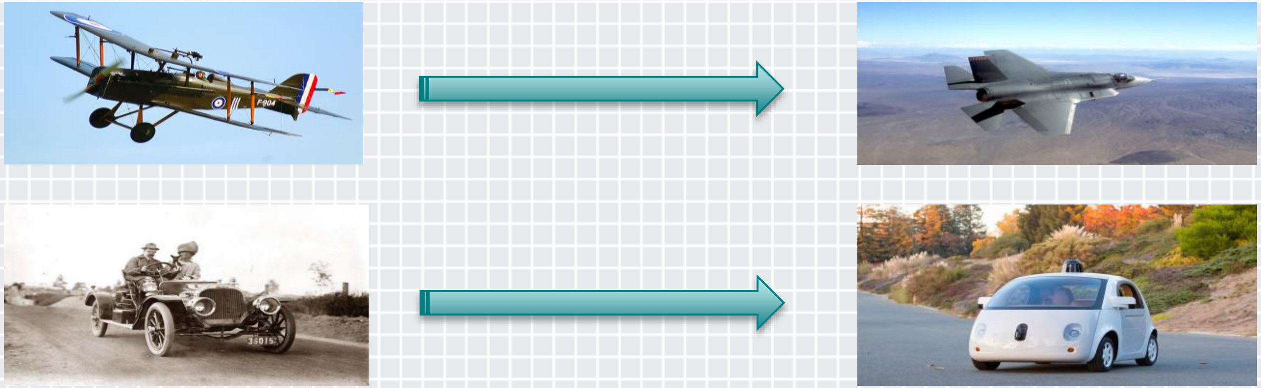
Risk-Based Security and Self-Protection

- Ongoing, automatic self-learning fraud detection model
- Real-time decisions/actions based on risk scores

- Based on an online-learning Bayesian engine
- Utilizing confidence based feedback loops
- Ranking based on constant case ratio



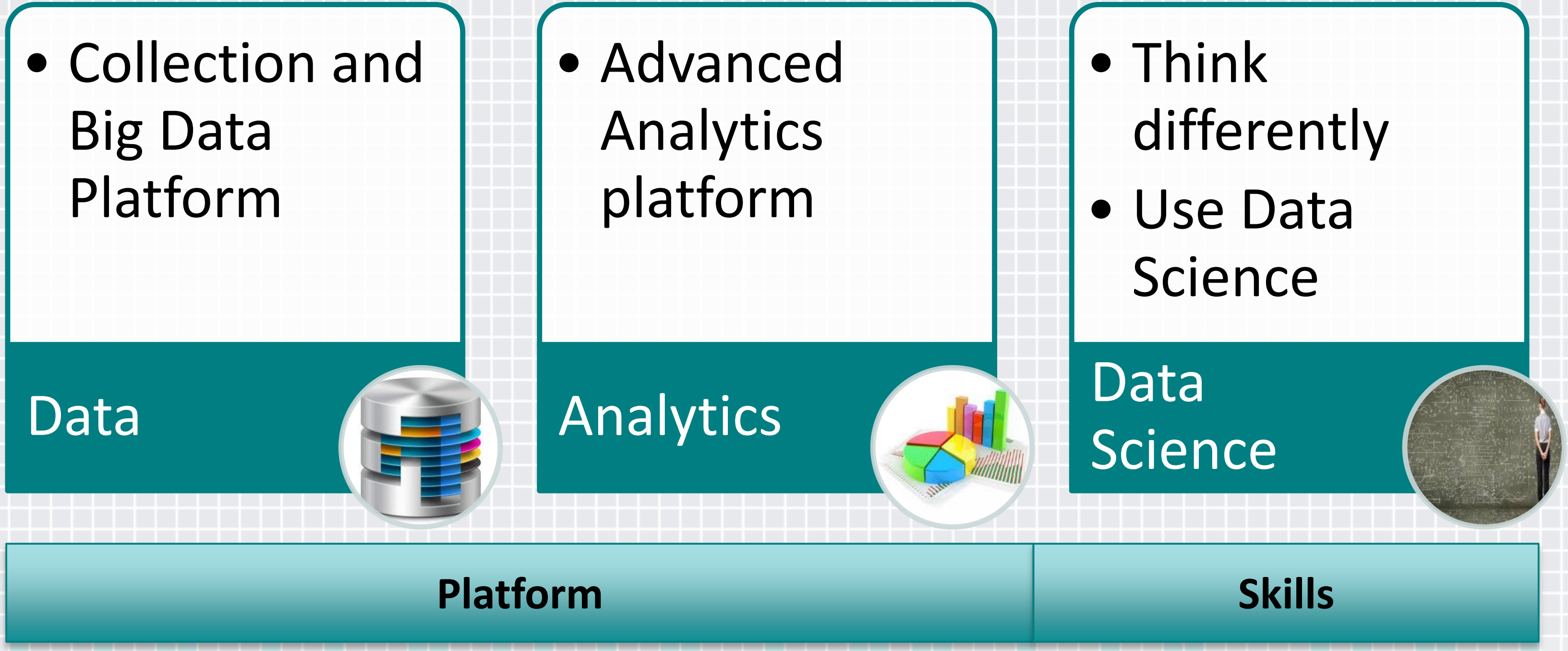
We Need to Transform…



So ***what*** **is needed to get there?**

So *How* do you get there?

The Transformation Journey



- Collection and Big Data Platform

Data

- Advanced Analytics platform

Analytics

- Think differently
- Use Data Science

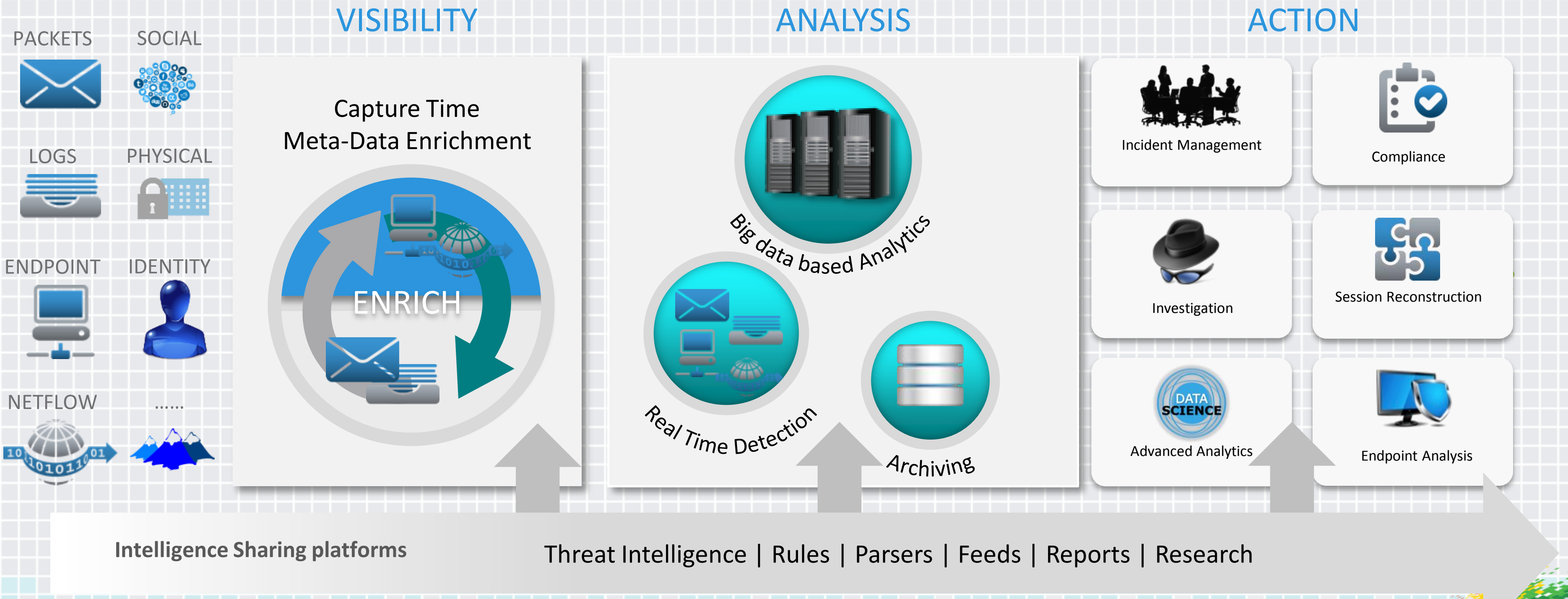
Data Science

Platform

Skills

Conceptual Security Analytics Platform

Capture, enrich and analyze data from across your network



PACKETS

SOCIAL

LOGS

PHYSICAL

ENDPOINT

IDENTITY

NETFLOW

......

VISIBILITY

Capture Time
Meta-Data Enrichment

ENRICH

ANALYSIS

Big data based Analytics

Real Time Detection

Archiving

ACTION

Incident Management

Compliance

Investigation

Session Reconstruction

Advanced Analytics

Endpoint Analysis

Intelligence Sharing platforms

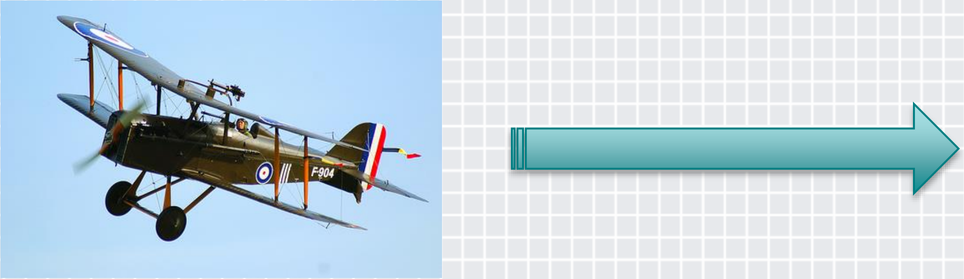
Threat Intelligence | Rules | Parsers | Feeds | Reports | Research

Taking the Data Science Path

- THINK DIFFERENTLY
  - Data science is not a feature -> it's a methodology!
  - Data science capabilities should synergize, empower and enhance the security experts -> not replace them!!

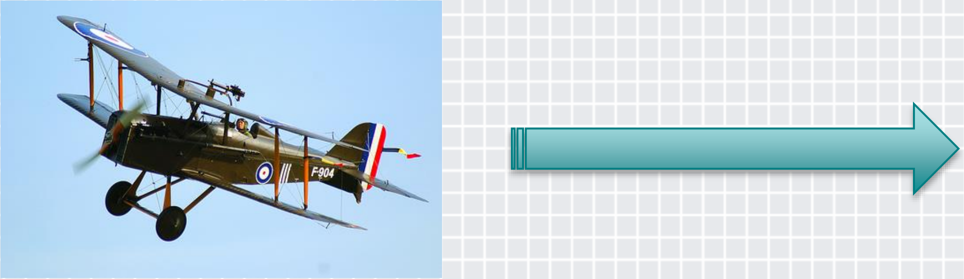
- USE DATA SCIENCE - Building your own (??)
  - Skills and expertise
  - How specific are your requirements
  - Learning from the past – BI, web analytics, retail analytics
  - Leveraging the community and crowd sourcing

Summary - Transforming Security Operations



- Current approaches are failing
- We have the data – take advantage of it, and leverage human intelligence for the intentions & decisions
- Security Analytics platforms with baked-in DS; Aggregate, prioritize, recommend, self-learning & crowd sourcing….

Summary - Transforming Security Operations



- Current approaches are failing
- We have the data – take advantage of it, and leverage human intelligence for the intentions & decisions
- Security Analytics platforms with baked-in DS; Aggregate, prioritize, recommend, self-learning & crowd sourcing….

WHY
VISION
WHAT
STRATEGY
HOW
TACTICS
dreamstime.com

CHANGE
Challenge today's security thinking

SESSION ID: STR-R04

Data Science Transforming Security Operations

Key message: Data science is a key methodology and technology in this transformation, its not merely a feature….

Alon Kaufman Ph.D. (alon.kaufman@rsa.com)

Director Data Science & Innovation, RSA

THANK YOU