Data Science in Cybersecurity

David Broggy,
Senior Security Consultant/Architect, Trustwave

TODAY’S SPEAKER

- David Broggy
- Senior Solutions Architect for Trustwave

- David has worked in cybersecurity since Y2K, travelling the world and delivering cybersecurity solutions to over 100 organizations, primarily Global 2000 and Fortune 500.

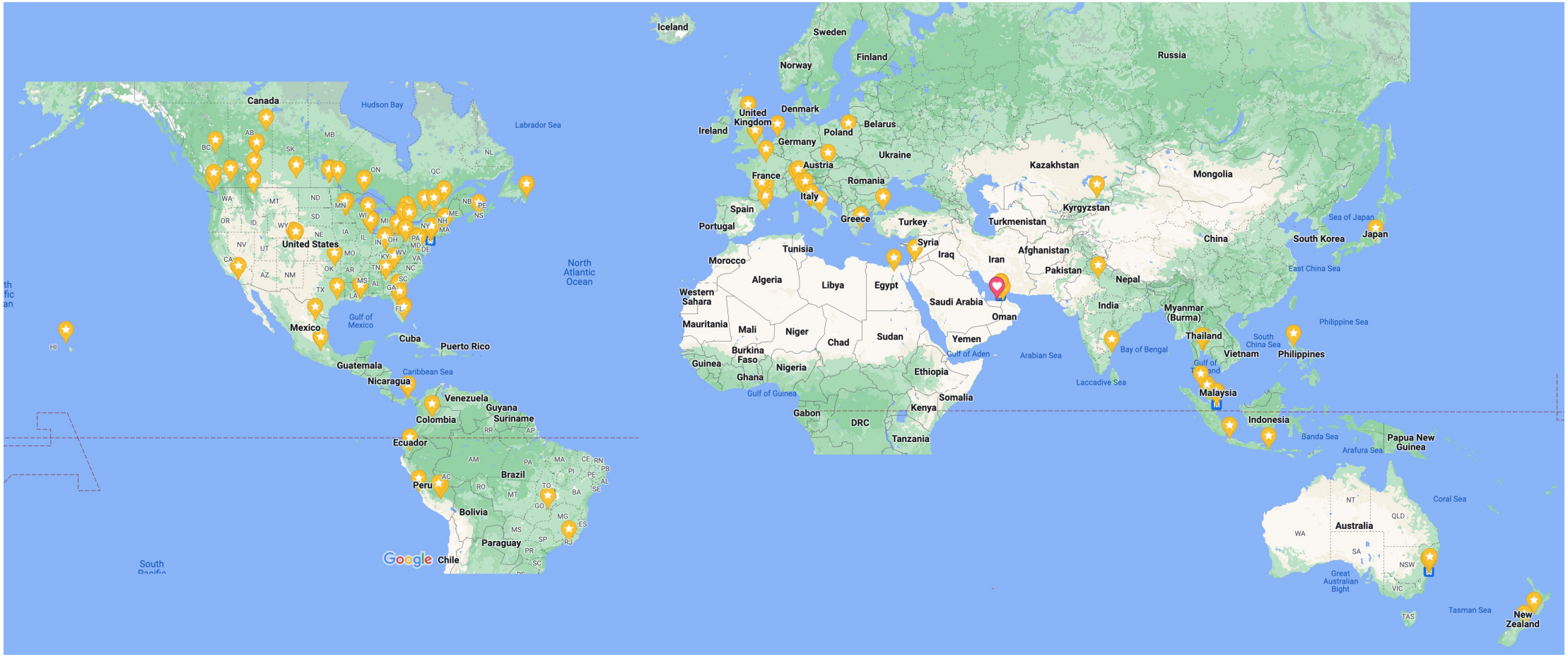
Some Things I Do

I work with the following cybersecurity categories/technologies (and more):

- SIEM – Splunk, QRadar, Azure Sentinel
- SOAR – XSOAR, Microsoft Logic Apps, Phantom
- EDR – Defender, Carbon Black, XDR, Cybereason, Crowdstrike
- Cloud – Azure/AWS/GCP
- Red/Blue/Purple team events – attack simulations.
- CASB
- CSPM
- DLP
- Information Protection
- IoT
- Zero Trust
- MITRE ATT&CK/SHIELD/D3FEND
- Threat Intelligence

Where have I worked in cybersecurity

Mostly large enterprises (over 100 unique companies)



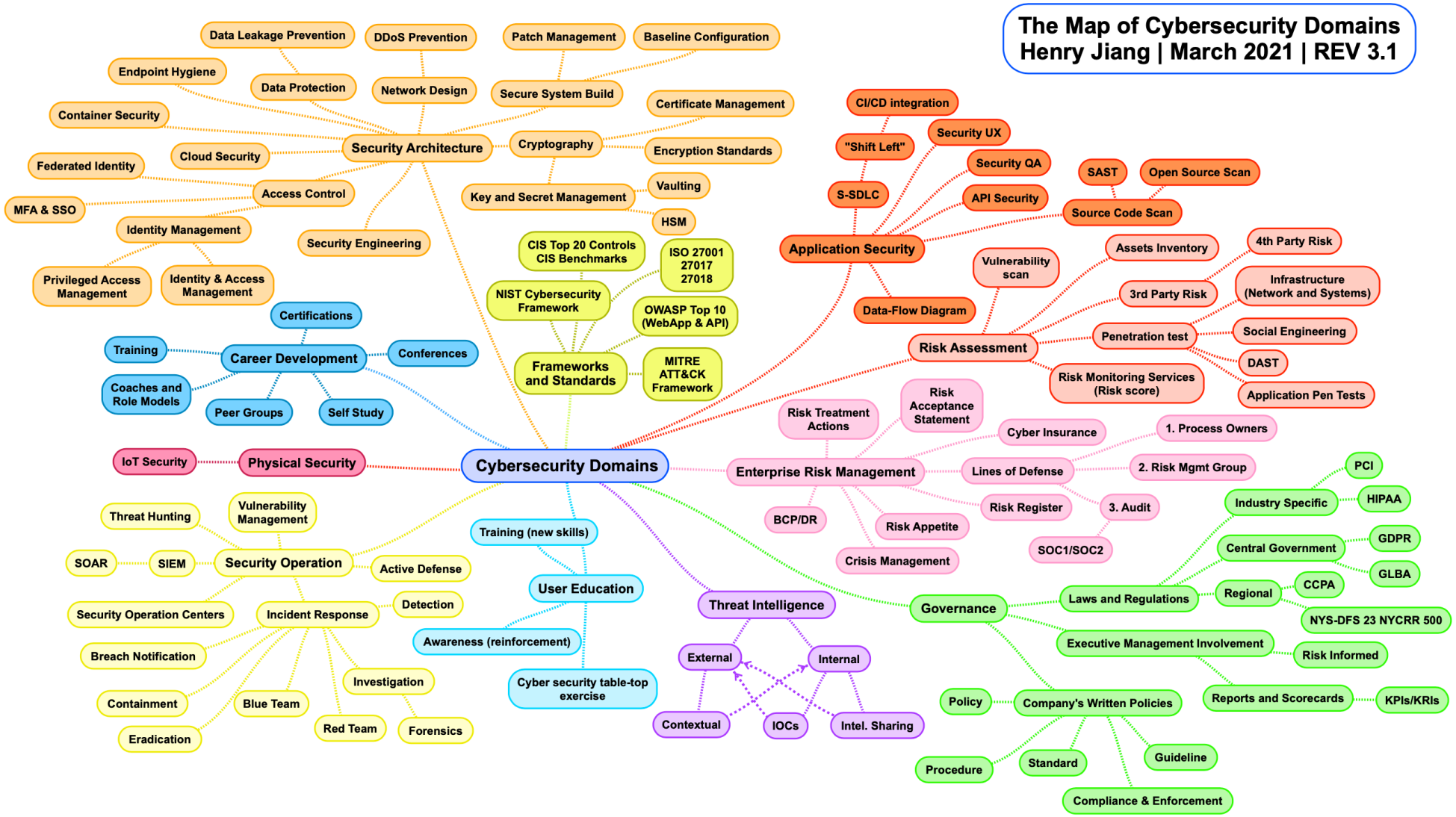
Objectives

- What is Cybersecurity
- Related tools and technologies
- How data science is used in Cybersecurity
- How ‘Cloud’ has been a benefit to data science
- Demonstration of cyber related data modeling tools like Splunk and Sentinel
- Core components (“Building Blocks”) for data modeling in Cybersecurity architectures
- Roles for data scientists in Cybersecurity
- Data Modeling tools in the Cloud
- Demo of Splunk and Sentinel after presentation

What is Cybersecurity?

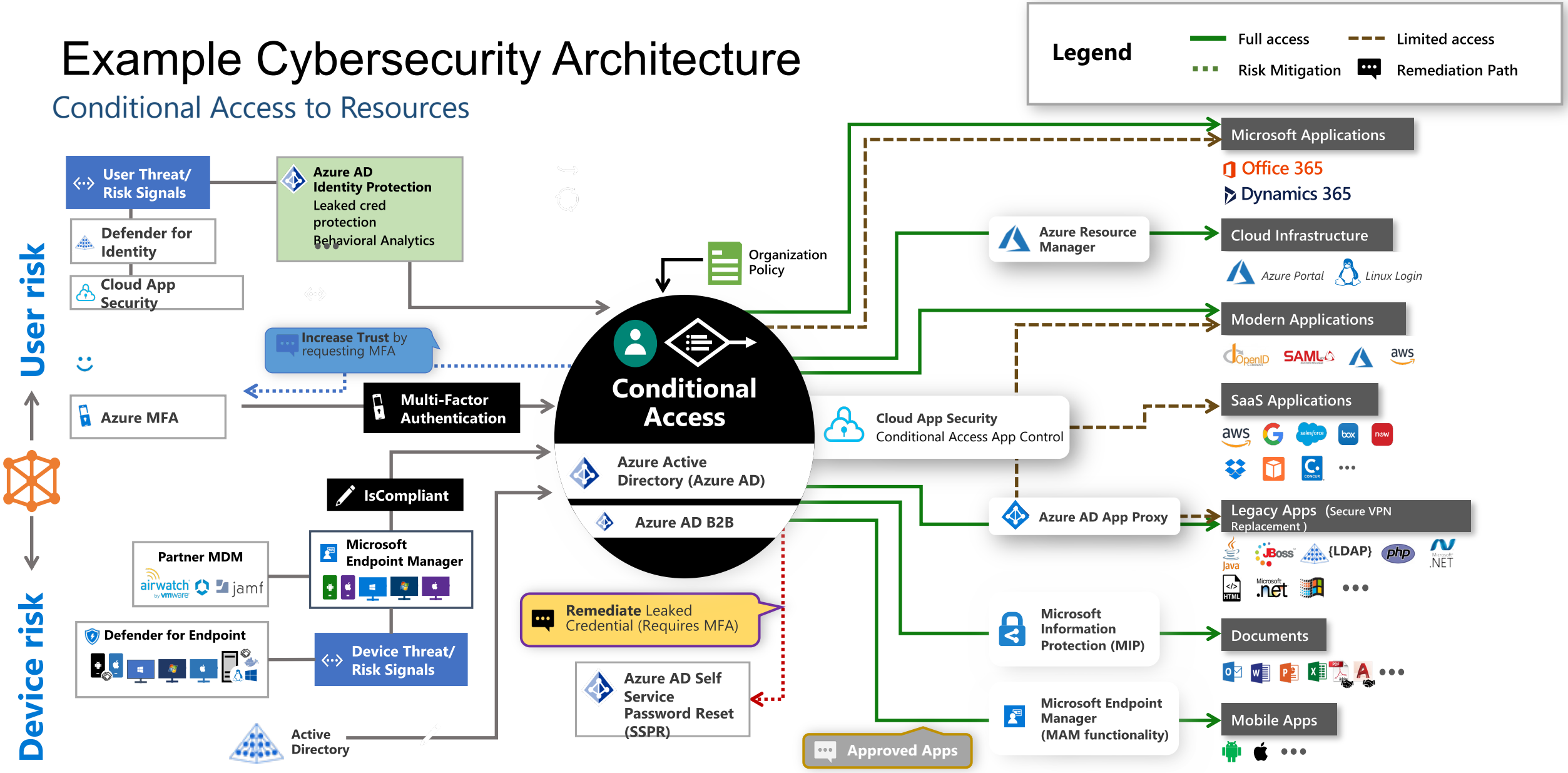
- Defense against the dark arts as they relate to IT
- Comprised of several ‘domains’ eg:
  - architecture
  - identity and access
  - network security
  - risk
  - assets
  - software
  - operations
  - testing

Cybersecurity Data Sources – by Security Domain

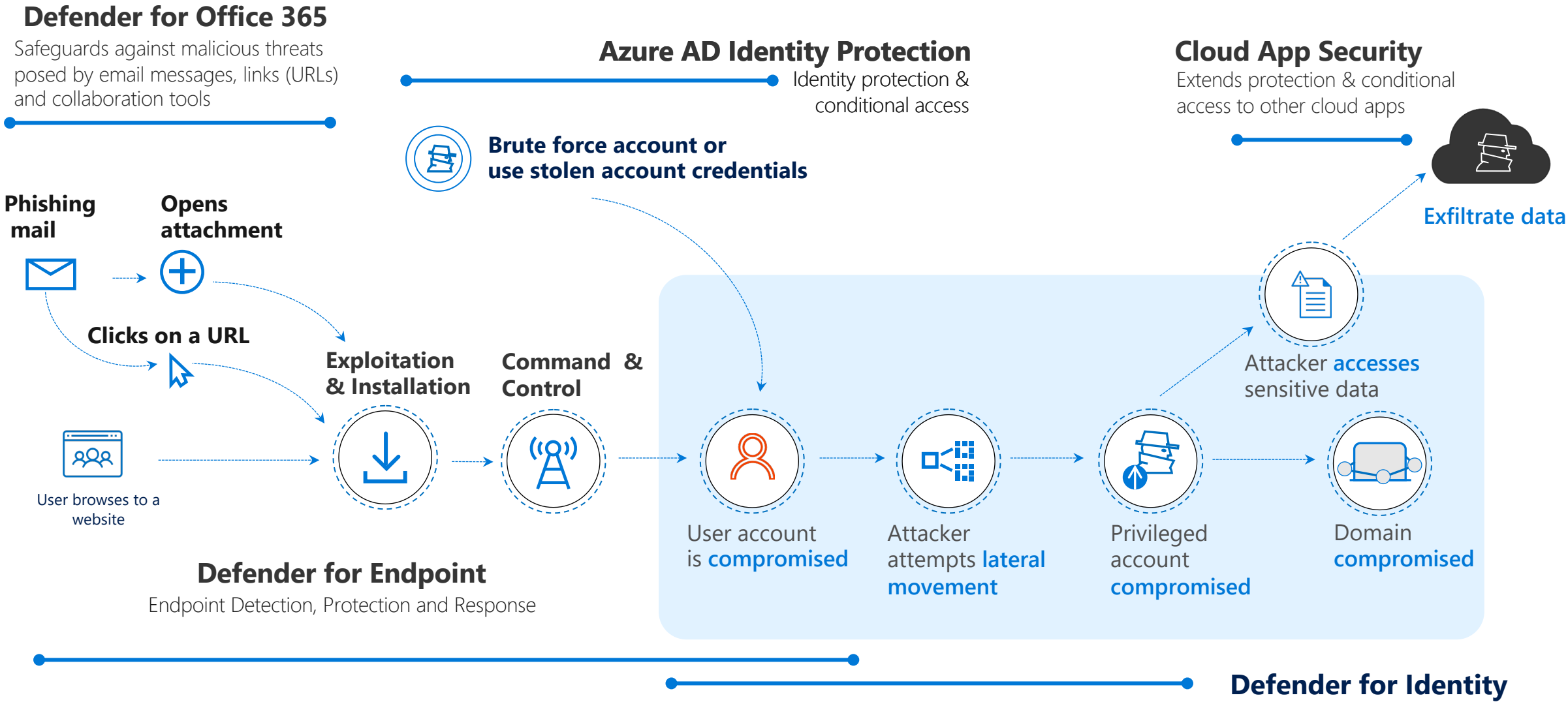


Example Cybersecurity Architecture

Conditional Access to Resources



EXAMPLE KILL CHAIN DETECTED USING CLOUD TOOLS



How is data science used in Cybersecurity?

Cybersecurity is ALL about data

- Collecting log activity from all network devices
- Identifying threats against those network devices, resources and users.

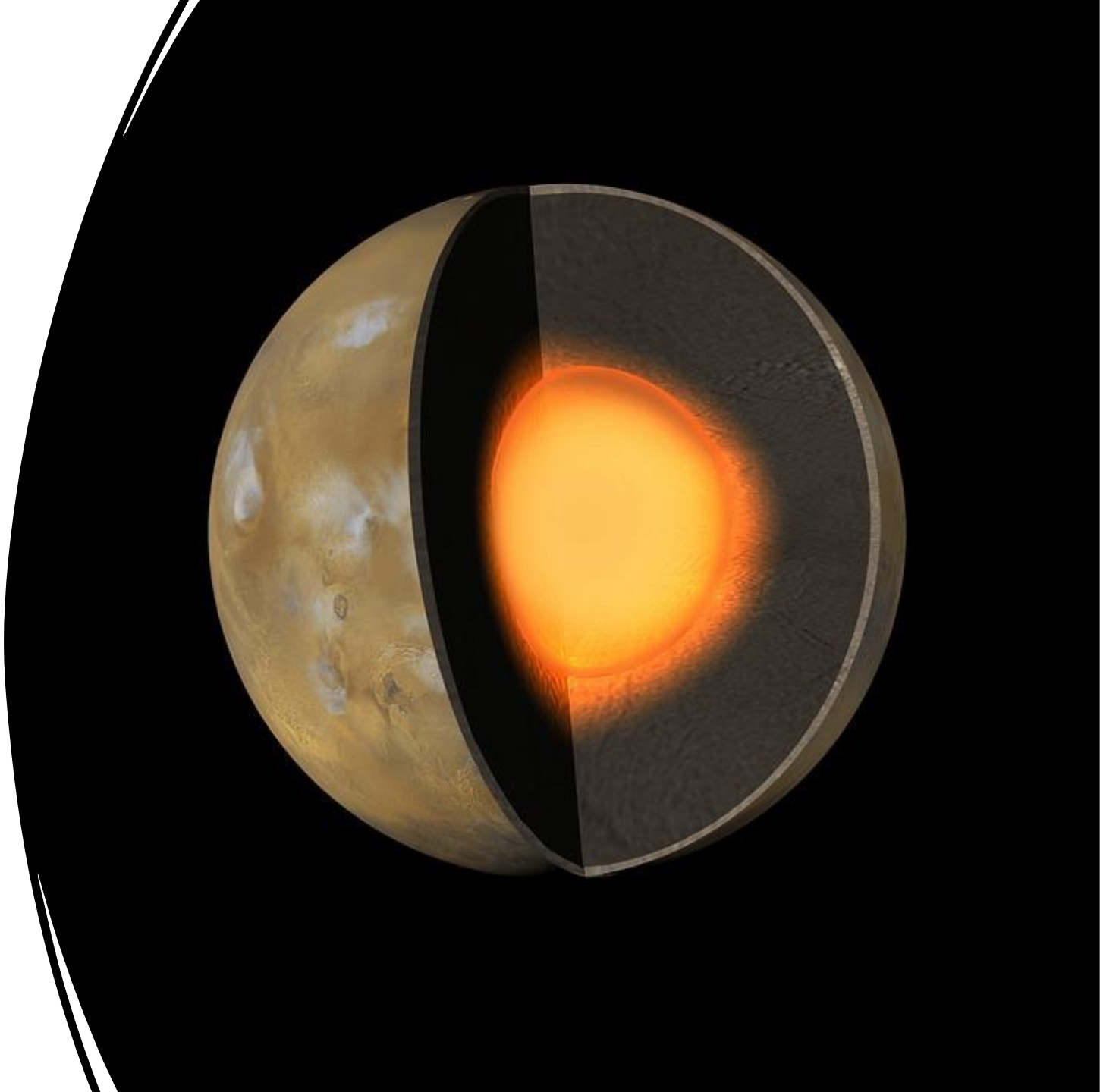
Hackers are natural data scientists

(or vice versa?)

- What are the core skills of a hacker?
  - Adept at searching through mountains of information
  - Skilled at creating tools that don't exist to find more data

Data Modeling Core Components for Cybersecurity Architectures

- data sources
- data normalization and modeling methods
- visualization tools
- efficient query methods



Data Set Structures and data model examples used in Cybersecurity

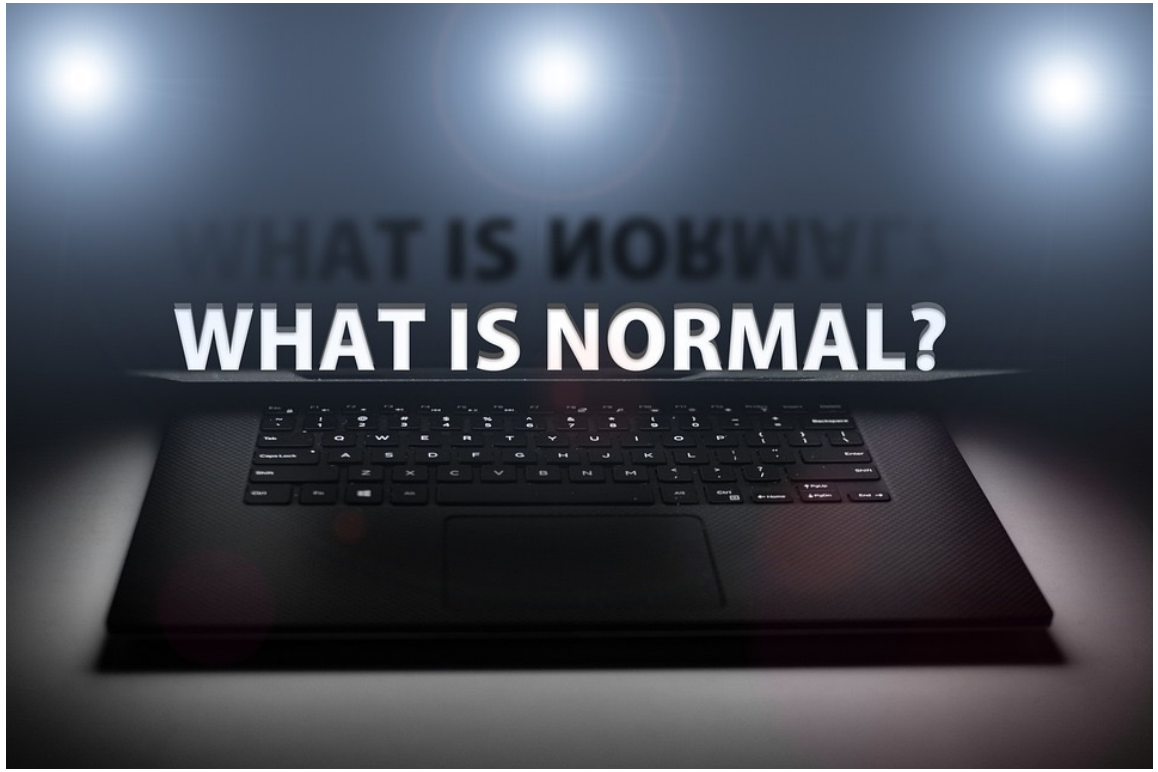
How are these logs structured and acquired?

- Formats:
  - Cybersecurity hates unstructured logs – no web scraping and multi line formats please.
  - JSON, CSV, delimited
- Acquisition Methods
  - Syslog – common for network appliances and unix servers.
  - API – Almost everything in the cloud has an API. Many modern applications support an API.

Why ‘Normalization’ of Data is a big thing

- https://docs.microsoft.com/en-us/azure/sentinel/normalization

- Microsoft and other vendors appreciate the challenges of working with the vast formats that data comes in.
- As such several ‘common log formats have been developed’ – eg. CIM, ASIM
- The advantages of using a normalized information model:
  - Queries get easier
  - Machine learning gets easier:
    - All your data belongs to us – in a cloud based data warehouse, data lake etc.
    - Common log format are normalized more automagically
    - Thus, less work for data scientist to prepare the data before working with it.



Common languages and formats used in Security tools

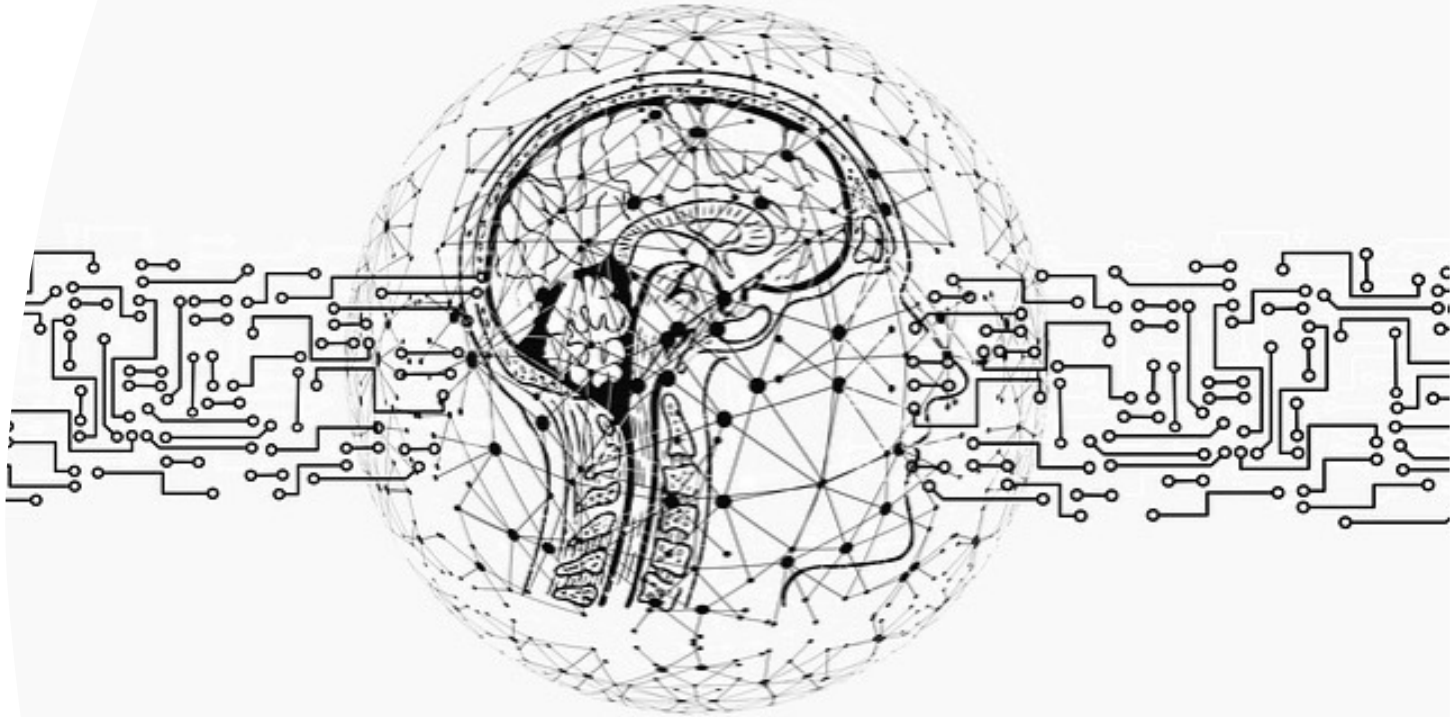
- Query languages:
  - SQL and variants (SPL – Splunk, AQL – QRadar, KQL – Microsoft)
- File formats:
  - JSON
  - CSV
  - syslog/CEF
- Programming languages: python, powershell, Visual Studio (and many many more!)

Examples of data sets in Cybersecurity

- Threat Intelligence – lists of evil IP addresses, domains, etc.
- Vulnerability references – what just attacked me? https://cve.org
- Logs, logs, logs – collected from everything on your network.

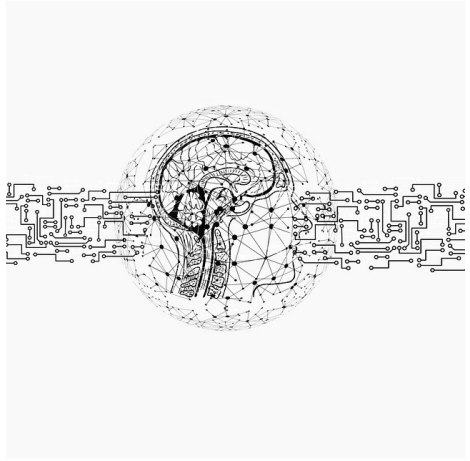
THREAT INTELLIGENCE

- Very important data sets of known bad stuff.
- Used by SIEM, EDR, CASB, WAF, Proxy, Email, etc.
  - Can help identify KBAs – Known Bad Actors
  - Example values include:
    - email addresses
    - IP addresses
    - Domain URLs
- OSINT – Open-Source Intelligence
  - Formerly thought of as free threat feeds but includes any open-source tools used to collect threat intelligence.
- Commercial Threat Intelligence tools: Recorded Future, VirusTotal



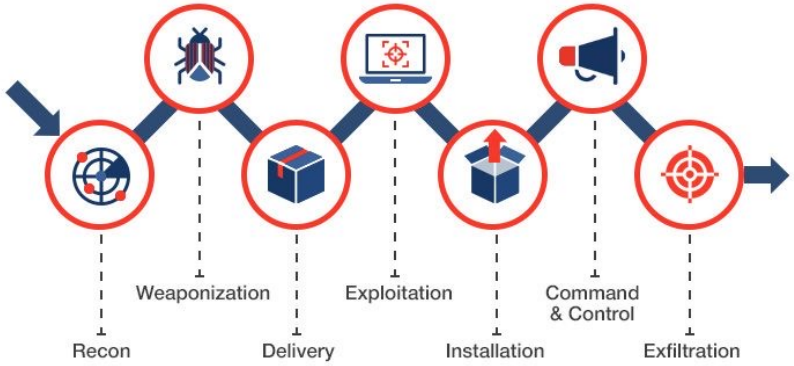
OSINT

- Open Source Intelligence
- **Several Categories, eg:**
  - **Government**
  - **IT/Internet**
  - **Law Enforcement**
  - **Geo**
  - **Media**
  - **Photos, videos, meta data (data which represents other data in some way)**



MITRE SECURITY FRAMEWORK

- What is the MITRE Security Framework?
- An ‘Attack Kill Chain’ encyclopedia
- A matrix (data set) of tactics and techniques used by threat hunters, red teamers, and defenders to better classify attacks and assess an organization's risk.



Reconnaissance 10 techniques	Resource Development 7 techniques	Initial Access 9 techniques	Execution 12 techniques	Persistence 19 techniques	Privilege Escalation 13 techniques
Active Scanning (0/2)	Acquire Infrastructure (0/6)	Drive-by Compromise	Command and Scripting Interpreter (0/8)	Account Manipulation (0/4)	Abuse Elevation Control Mechanism (0/4)
Gather Victim Host Information (0/4)	Compromise Accounts (0/2)	Exploit Public-Facing Application		BITS Jobs	Access Token Manipulation (0/5)
Gather Victim Identity Information (0/3)	Compromise Infrastructure (0/6)	External Remote Services	Container Administration Command	Boot or Logon Autostart Execution (0/14)	Boot or Logon Autostart Execution (0/14)
Gather Victim Network Information (0/6)	Develop Capabilities (0/4)	Hardware Additions	Deploy Container	Boot or Logon Initialization Scripts (0/5)	Boot or Logon Initialization Scripts (0/5)
Gather Victim Org Information (0/4)	Establish Accounts (0/2)	Phishing (0/3)	Exploitation for Client Execution	Browser Extensions	Create or Modify
Phishing for Information (0/3)	Obtain	Replication	Inter-Process Communication (0/2)		

Situational Awareness Dashboards

- The Basic Question: How is a ‘data set’ affecting my resource?
- The Example: Russia attacking critical government infrastructure via cyberspace.
- The data sets:
  - OSINT
  - Paid-for intelligence feeds – blacklisted IPs, domains, etc.
  - internal assets and related network traffic
- The solution: Create visual dashboards to provide an ‘at a glance’ view of all Russian related activity against my ‘data sets’ (network traffic logs, etc).

CLOUD SECURITY TOPICS

- Cloud has changed the approach to cybersecurity
- **Cloud provides huge centralized data stores and tools to query and protect it.**

Data Sources in Cybersecurity

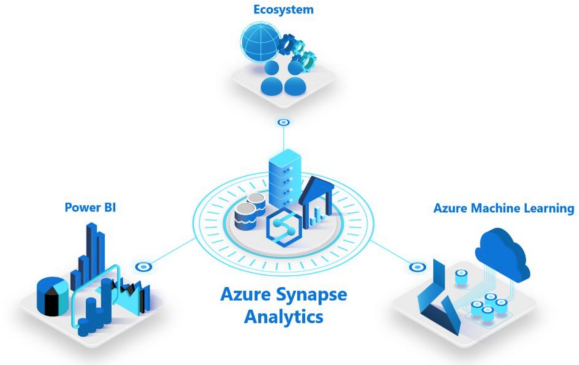
- There are SO MANY data sources in security.
- Each source may have a different log format to parse
- Comparing one log source to another is key. (correlations)
- The cloud has a lot to do with the evolution of better cybersecurity.

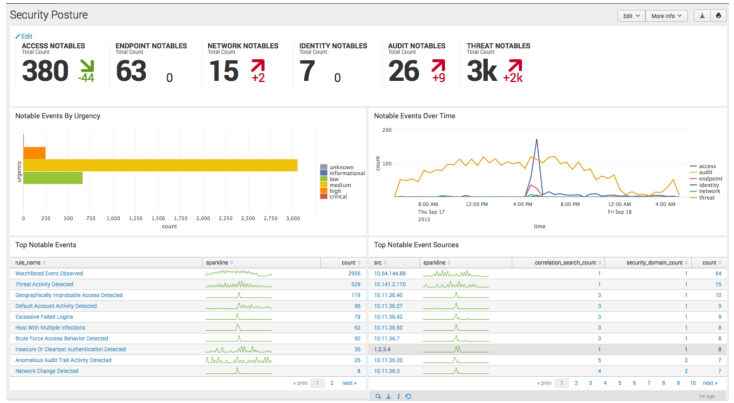
MODERN SECURITY DEFENSES (TOOLING & ARCHITECTURE)

- SIEM
- SOAR
- EDR/XDR
- Threat Intelligence
- CSPM
- Cloud Security Analysis Tools – Sentinel and Splunk

Data Model Tools in the Cloud

- Azure Data Factory – no-code data processing
- Azure Synapse – machine learning
- Jupyter Notebooks - https://jupyter.org/
  - Python analytic library
  - MsticPy - enrich the data with Threat Intelligence, geolocations and Azure resource data
  - https://github.com/microsoft/msticpy
- Splunk Cloud – for much more than cybersecurity!





Why Use a tool like Splunk over Excel

- Excel has no ability to ’accelerate’ data
- Splunk allows you to customize data models which improves indexing, resulting in lightning-fast queries over billions of rows of data.
- Splunk has tons of tools to acquire, process and present data
- Splunk support machine learning

Roles for Data Scientists in Cybersecurity

- Job posting example: indeed.com
  - Security Engineer – Machine Learning and Cybersecurity Analyst
    - $95-$200k/year (US dollars)
    - 4 years cybersecurity experience
    - 3 years Splunk experience
- Recommended skills in general:
  - Data Science degree
  - Understanding of Cybersecurity – through schooling or certifications
  - Basic to deep programming skills is a bonus.

- Tons of low cost and free tools: eg. portal.azure.com (cloud), splunk.com (data modeling).

Questions?

Final Words

- Data is everywhere – seek out and open your mind to the possibilities.
- Data is what gives corporations their value
  - ‘keys to the kingdom’ – common hacking endgame.
- CIO/CISOs are constantly in search for meaning in their own data.
- Think like OSINT – don’t always limit yourself to a single data set.
- Don’t stop learning. Think like a leader. Work as a team.

Demos

- Splunk, Data Models
- Azure Sentinel

Azure Sentinel

Data Modeling and Cybersecurity Resources

- Splunk & Machine Learning – Data Modeling discussion
  - https://www.youtube.com/watch?v=N3FL6rawDLQ&ab_channel=Splunk%26MachineLearning
- Udemy.com – cheap training stuff
  - https://www.udemy.com/courses/search/?src=ukw&q=%22data+modeling%22
- OSINT – Open Source Intelligence
  - https://www.nixu.com/blog/open-source-intelligence-its-incredible-what-you-can-find-public-sources#:~:text=OSINT%20is%20one%20of%20many,SIGINT%20can%20overlap%20with%20OSINT
  - https://itsec.group/blog-post-osint-guide-part-1.html
  - https://en.wikipedia.org/wiki/Open-source_intelligence
  - https://owasp.org/www-chapter-ghana/assets/slides/OWASP_OSINT_Presentation.pdf
  - https://www.csoonline.com/article/3445357/what-is-osint-top-open-source-intelligence-tools.html
  - https://www.secjuice.com/introduction-to-open-source-intelligence-osint/
- Azure Data Factory
  - https://docs.microsoft.com/en-us/azure/data-factory/introduction
  - https://docs.microsoft.com/en-us/azure/data-factory/solution-templates-introduction

Cybersecurity Recommended Reading

- Cyber Strategy: Risk-Driven Security and Resiliency
- Hacking Exposed 7: Network Security Secrets and Solutions
- Hacking: The Art of Exploitation, 2nd Edition
- Hands on Hacking: Become an Expert at Next Gen Penetration Testing and Purple Teaming
- How to Hack Like a GHOST: A detailed account of a breach to remember (Hacking the planet Book 8)
- How to Hack Like a GOD: Master the secrets of hacking through real life scenarios (Hacking the planet Book 2)
- How to Hack Like a LEGEND: A hacker’s tale breaking into a secretive offshore company (Hacking the planet Book 7)
- Hunting Cyber Criminals: A Hacker's Guide to Online Intelligence Gathering Tools and Techniques
- Learn Azure Sentinel: Integrate Azure security with artificial intelligence to build secure cloud systems
- Linux Basics for Hackers: Getting Started with Networking, Scripting, and Security in Kali
- Metasploit: The Penetration Tester's Guide
- Microsoft 365 Compliance: A Practical Guide to Managing Risk
- Microsoft Azure Security Infrastructure (IT Best Practices - Microsoft Press)
- Microsoft Azure Sentinel: Planning and implementing Microsoft’s cloud-native SIEM solution (IT Best Practices - Microsoft Press)
- PTFM: Purple Team Field Manual
- Pentesting Azure Applications: The Definitive Guide to Testing and Securing Deployments
- Red Team Development and Operations: A practical guide
- The Hacker Playbook 3: Practical Guide To Penetration Testing
- The Web Application Hacker's Handbook: Finding and Exploiting Security Flaws
- Web Application Defender's Cookbook: Battling Hackers and Protecting Users
- Zero Trust Security: An Enterprise Guide

Thank you!