Leveraging Disparate Enterprise Data for Cybersecurity Purposes

Introduction
A vast array of data can theoretically be useful for cybersecurity purposes:
- Yet it is rarely fully leveraged because the data is not clean, complete, integrated, or searchable.
- Creation of a cybersecurity data lake may seem like the answer, but the result of such efforts may not bring the intended results.
- By approaching the task from the perspective of the desired end results, efficient translation to business insights is achievable.

Methods
- Aim for a usable lake, above having a large lake
- Fully evaluate each potential source before ingestion
- Focus efforts on preparing the data for efficient analysis
- Consider early how to reduce volumes of scanned data, for cost and efficiency purposes

Results
- A cyber data lake poised to address the most pressing cybersecurity use cases expeditiously and efficiently
- A replicable process for augmenting the lake with additional future data sources
- Enhanced ability to interpret results accurately and translate them to business insights.
- Platform for integrating AI/ML models

FUSION TEAM
- Cybersecurity experience and certifications
- Network architecture and operations
- Software and script development
- Database administration and operations
- Expertise across multiple OS/platforms
- Source system familiarity
- Technical writing and documentation
- AI and Machine Learning
- Statistical Analysis of large dataset

APPROACH
1. Use Case Selection
2. Data Sources
3. Analytic Preparation
4. Data Evaluation

Start with a limited number of very specific use cases
Define integration methodology to ensure scalable and translatable analytics
Deeply analyze data content, breadth, and scale across selected sources
Identify and vet possible data sources within your organization

Unique Capabilities
- Extensible architecture designed to facilitate additional data sources and further functionality
- Incorporates the organization’s custom business rules (e.g., NAC, VLANs, etc.) to elucidate behaviors of source systems
- Provides scalability, but also history and granularity (not just a bigger hammer)
- Enables data correlation across disparate retention periods

Multiple cybersecurity uses:
- Hunt
- Investigate
- Model
- Predict
- Validate
- Monitor
- Multiple analytic use cases:
  - Data Exfiltration
  - LRCs
  - Beaconsing
  - Anomaly detection, etc.

Recommendations
- Recognize the full implications of building in the cloud
- Paradigm shift
- “Cattle, not pets”
- Implications are often obscure

Conclusion
It is possible to create a large-scale, cloud-based data lake that is sustainably positioned for data modeling, based on careful preparation and thorough data evaluation.