ENLIGHTEN

# Cyber Defense at Machine Speed

**Duane Shugars**
SVP, General Manager
dshugars@eitccorp.com

**John Eubank**
Director, Client Engagement
jeubank@eitccorp.com

991 Corporate Blvd #350,
Linthicum Heights, MD 21090
Phone: +1.410.449.1201

## ABSTRACT

Protecting U.S. cyber assets is a top-level priority in the defense and intelligence communities. To defend the Department of Defense's (DoD) information networks, cyber analysts must comb through the vast, unstructured volume of cyber defense data to detect, assess, and mitigate cyber threats and act quickly. Faster and more flexible solutions are needed to keep pace with the evolving cyber terrain and ever-increasing amounts and types of data. This dynamic and expanding environment demands that defensive cyber operations run at machine speed— not human speed. In addition to speed and flexibility, security, portability, and accessibility must be factored into any viable solution. And, of course, cost. Enlighten IT Consulting (Enlighten), LLC is meeting the challenges of the volume and complexity of big data by applying artificial intelligence (Ai) and machine learning capabilities. These solutions including: the Cyber Learning Engine (CLE); RAPID, a comprehensive analytic solution; and the Rapid Analytic Deployment and Management Framework (RADMF®).


### AUTOMATE IDENTIFICATION OF CYBER THREATS WITH THE CYBER LEARNING ENGINE (CLE)

As the volume of DoD cyber data grows exponentially, identifying threats to networks becomes an increasingly complex task. Searching for known threats — such as those identified in existing open source and commercial feeds — can be effective but is hampered by manual processes and procedures. This method does nothing to identify new threats (i.e., those not previously identified by other sources or feeds).

To address this problem, Enlighten's data science team set out to create a model that could learn about the nature of threats from a variety of data sources and use this understanding to accurately predict the probability of new indicators being malicious.

The resulting CLE (Figure 1) is trained to identify previously unknown threats to a network, automatically, without human supervision. This engine is based on a model that combines operational user feedback with 150 diverse global datasets and an infrastructure that can scale with on-demand, high-density compute.

This CLE was developed based on a deep learning (DL) concept of using artificial neural networks to process data in layers. As more data is entered, the model's conclusions about bad actors in the network become more accurate over time. For example, the new CLE progressed from a 75% probability of correctly identifying malicious content up to 95% accuracy. Even more significant, it identified signs of a network breach, also known as Indicators of Compromise (IOC), not previously reported in any open or non-open source data and later identified as threats.

DL models that can search through terabytes of data and reach accurate conclusions without human intervention are essential for future success. For Enlighten, the key will be to continue to feed this model with as much data as possible. The goal is to develop a model that can independently determine operational normal across the DoD, allowing it to more easily predict and prevent cyber threats.

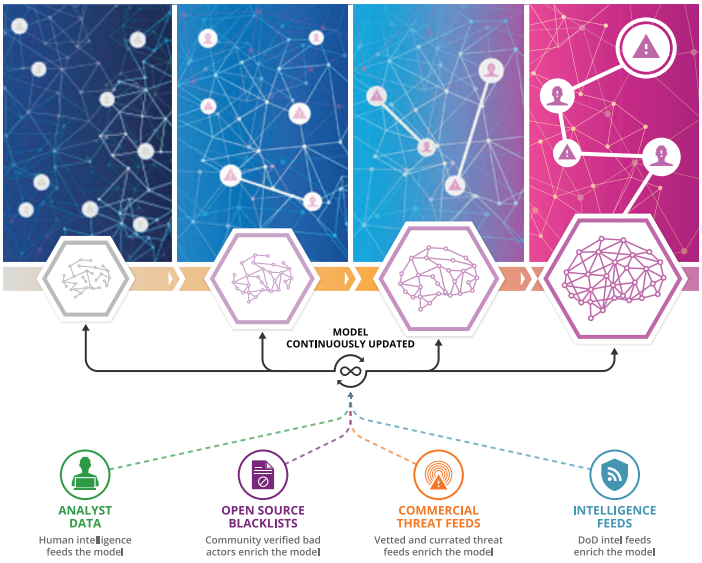



FIGURE 1: *As data is added and the model learns more, its predictions about bad actors in the network become more accurate.*

### QUICKLY DEVELOP AND DEPLOY ANALYTIC SOLUTIONS WITH RAPID

In tandem with CLE development, Enlighten developed RAPID as a comprehensive analytic solution that allows big data scientists to extract valuable insights from DoD information network data. This solution utilizes common languages and interfaces — including

### CHOOSE THE RIGHT INFRASTRUCTURE

At the foundation of all successful Ai lies a mature infrastructure that enables computing, data storage, and analytics for effective modeling. Enlighten supported the development of the Big Data Platform (BDP), a robust and scalable cloud-based architecture capable of ingesting, storing, and visualizing multiple petabytes of cyber data. Its distributed data structures and streaming ingest capabilities provide storage and retrieval rates in the millions of records per second. Enlighten also developed and deployed a suite of cyber situational awareness analytics to the BDP, giving analysts tools for accelerated attack detection, diagnosis, and threat mitigation.

Initially developed in 2012 for the Defense Information System Agency (DISA), the BDP is currently used by an active, diverse, and operational base of mission partners across the DoD. The platform has the power to handle the ingest volume and velocity of critical datasets and provides a standard analytic development interface to support the DoD in pioneering Ai and DL capabilities for cyber defense.

Structured Query Language (SQL), Python, and Apache Spark — to perform machine learning used in the CLE.

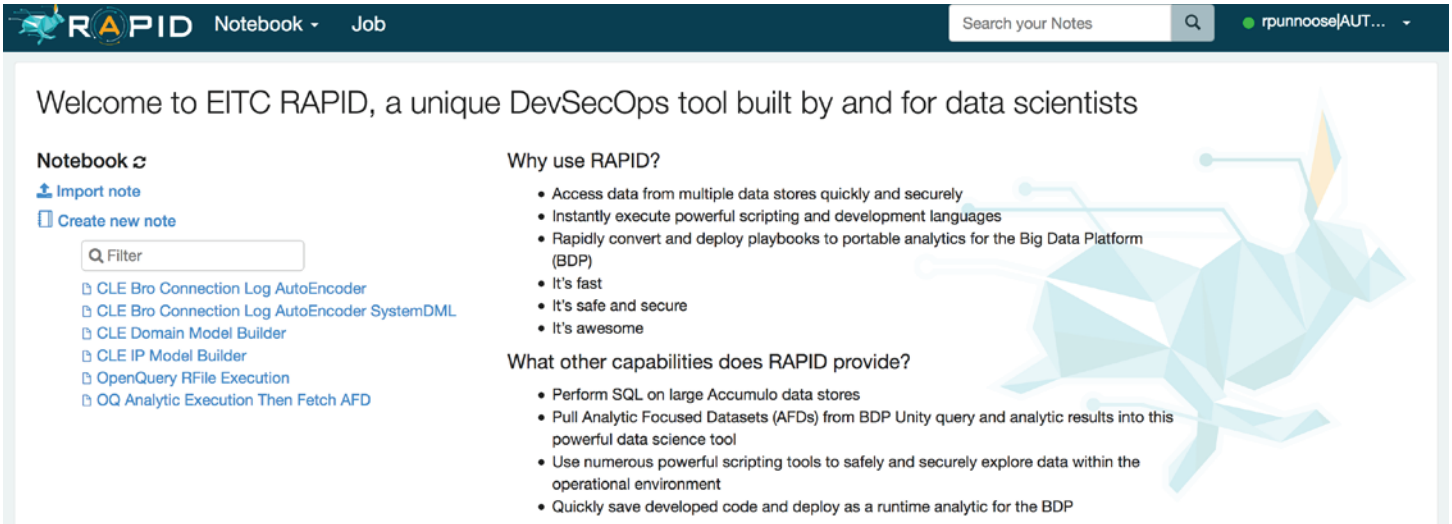
Similar analyst tools have been developed for smaller datasets, but RAPID — scalable to effectively handle petabytes of data — fills a critical gap for data scientists. During prototype testing, Enlighten's data science team found that RAPID could handle billions of events across multiple feature sets and that immediate deployment of analytics on a big data platform does not have to be done in isolation.

RAPID allows data scientists to collaborate in the same web-based environment and provides analysts with a notebook to test, develop, scale, and deploy with quick turnaround. (Figure 2) It also contains multiple security layers to limit individual access to different sets of DoD data, tightly controlling security information at the cell level.

### DEPLOY A POWERFUL, COMPREHENSIVE BIG DATA INFRASTRUCTURE USING RADMF

To keep pace with adversaries who constantly change attack vectors and methods, Enlighten recognized the need to provide analytic developers with a sophisticated framework to rapidly develop and test their analytics at a reasonable cost. Enlighten developed the Rapid Analytic Deployment and Management Framework (RADMF) for the key analytic components of the BDP (Figure 3).

FIGURE 2: *RAPID's notebook supports collaboration and quick turnaround.*



Within minutes, RADMF deploys and sets up a BDP environment with more than 40 open source technologies to support analytics development at production scale. Using Amazon™ Web Services (AWS) rather than a traditional computer network eliminates the need for hardware, power, space, cooling and sustainment— and negates the cost of hiring staff to monitor and maintain a data center. RADMF can be turned on and off as needed to control costs.

In RADMF, government customers are developing, testing, and validating analytics; ingesting and visualizing data; and performing computations and algorithms. RADMF has also proven to be an excellent training environment, as new analytics are developed and released. It provides an instant feedback loop from the analyst to the development team as they iterate through the development process.

The platform's advanced cyber mission tools also include inline click-to-content for packet capture (PCAP), malware analysis, secure DevOps of analytics, machine learning, and Ai capabilities to assist a variety of use cases that include intelligence correlation and fusion.

The continuous back-end tech refresh means that developers are always working with the latest BDP release. The latest release now incorporates operator feedback received throughout several years from multiple DoD customers, with more advanced tools to rapidly and efficiently meet cybersecurity mission requirements.

## SUMMARY

While many industries are embracing Ai and DL, the DoD is just beginning to apply this technology to its complex cyber challenges. By providing DoD customers with easy-to-use solutions for deployment, scalability, and sustainment of a big data analytic environment, Enlighten frees them from concerns about the platform and technologies they are using, enabling them to focus on developing cost-effective analytics to meet and exceed mission objectives.

FIGURE 3: *Through the use of best-in-class open source technology, combined with scalable commercial cloud infrastructure, RADMF provides a comprehensive and accredited industry solution for many long-standing problems in the cyber and big data domains.*

| PROBLEM | RADMF SOLUTION |
|---|---|
| Accessibility | Makes data accessible and usable for DoD mission partners through a new distributed query (i.e., query data across all partner platforms from a single access point) as well as shared analytics and workflows |
| Cost | Eliminates need for stove-piped and point solutions with large license fees. Along with the BDP and commercial cloud deployment, reduces Total Cost of Ownership compared to on-premise solutions |
| Flexibility | Ingests petabytes of data in any format; performs data normalization/transformation to assist analysts |
| Portability | Deploys to enterprise, tactical & regional environments |
| Security | DoD cloud-compliant environment ensures data security down to cell level across all tiers, which is critical for customer downstream-sharing of findings across titles of authority and titles of command |
| Speed | Rapidly deploys the full platform in four, simple, user interface-driven clicks— or by voice command |

ENLIGHTEN

991 Corporate Blvd #350,
Linthicum Heights, MD 21090
Phone: 410.449.1201
Fax: 410.255.5522