HACSAW: A Trusted Framework for Cyber Situational Awareness

Leslie C. Leonard, PhD
William J. Glodek
April 10, 2018

Outline

- **HPCMP Ecosystem**
- **Cyber Situational Awareness (SA) Initiative**
- **HACSAW – The Big Picture**
  - Data Repository
  - Development Workflow
- **Overview of Operational Use Cases**
- **Summary**
- **Questions**

HPCMP Ecosystem

Users

DOD HPC

A technology-led, innovation-focused program committed to extending HPC to address the DoD's most significant challenges

DoD Supercomputing Resource Centers (DSRCs)

- U.S. Air Force Research Laboratory DSRC
- U.S. Army Research Laboratory DSRC
- U.S. Army Engineer Research and Development Center DSRC
- Maui High Performance Computing Center DSRC
- US Navy DSRC

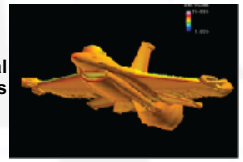
Networking and Security

Defense Research & Engineering Network (DREN)

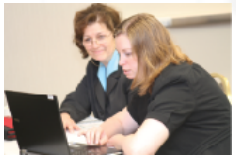
Computer Network Defense, Security R&D, and Security Integration

Software Applications

Core Software



Computational Environments

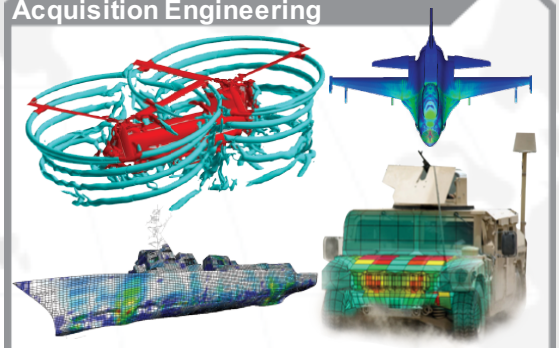


Education and Training

HPC User Support

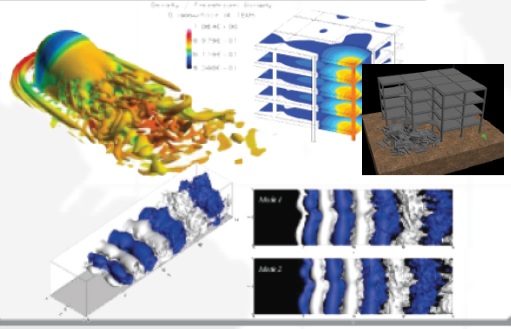
Results

Acquisition Engineering



Test and Evaluation

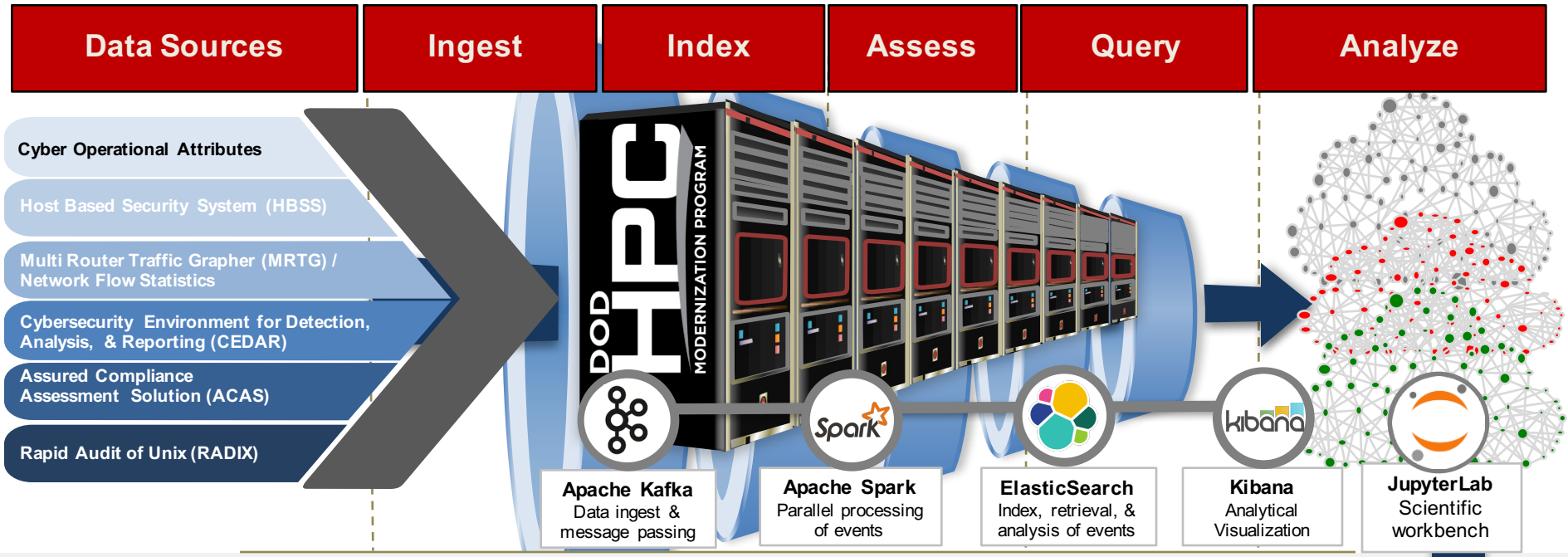
Science and Technology



Cyber Situational Awareness Initiative

- **Executive Steering Group "… examine the applicability of HPC to cyber situational awareness (SA)"**
- **HPCMP is well-positioned to leverage HPC systems to address complex cybersecurity problems**
  - World-class computational resources leveraged by the RDT&E community
  - Leading-edge software applications for computational analysis capabilities
  - National research and engineering network – DREN
- **Multi-disciplinary, multi-year project leveraging expertise from HPCMP (e.g., Security, Networking, Centers, Software Applications) and external collaborators**

HACSAW – The Big Picture

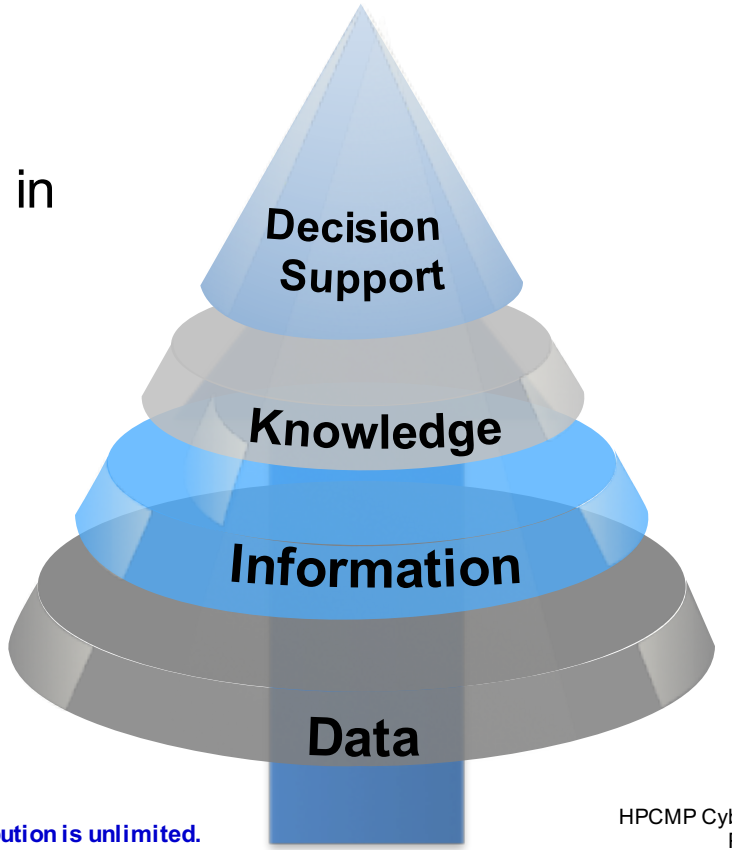


Cyber SA Mission Essential Tasks (MET)

Data Repository

Most comprehensive cybersecurity data set available to DoD R&D community

- ✓ Collection of data sources from Internet Access Points (IAPs) to regional Service Delivery Points (SDPs) to the host-level
- ✓ Non-anonymized data
- ✓ Contains operational attributes
- ✓ Rapid acceleration and exponential growth in size and complexity



HPC Development Workflow

1. DATA EXPLORATION

Identify relevant data sources and its underlying structure, purpose, and usefulness. In this stage, collaborators will exercise APIs and ontology to be used in the initial analytic development.

2. INITIAL DEVELOPMENT

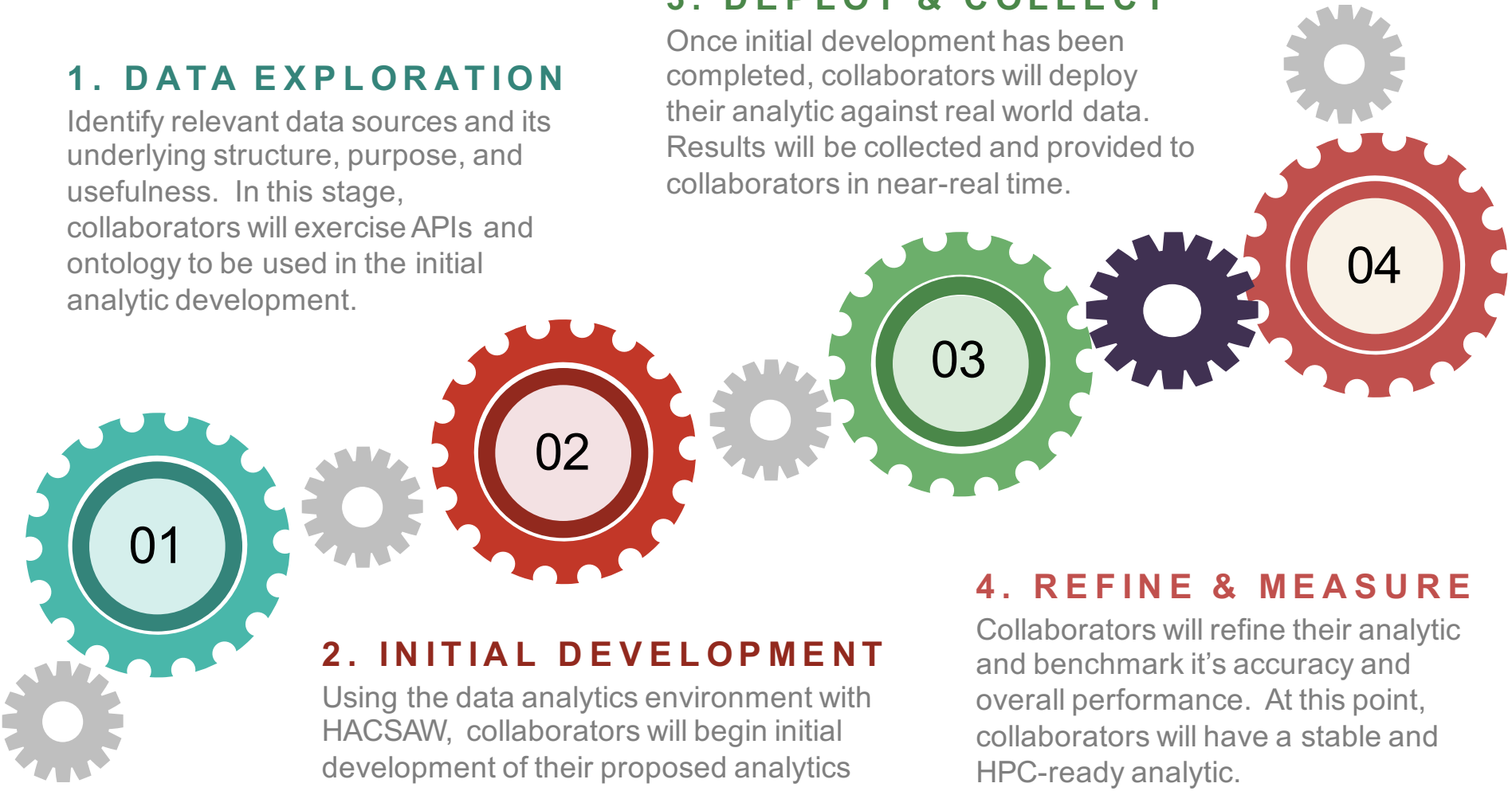
Using the data analytics environment with HACSAW, collaborators will begin initial development of their proposed analytics by working with real HPC data.

3. DEPLOY & COLLECT

Once initial development has been completed, collaborators will deploy their analytic against real world data. Results will be collected and provided to collaborators in near-real time.

4. REFINE & MEASURE

Collaborators will refine their analytic and benchmark it's accuracy and overall performance. At this point, collaborators will have a stable and HPC-ready analytic.



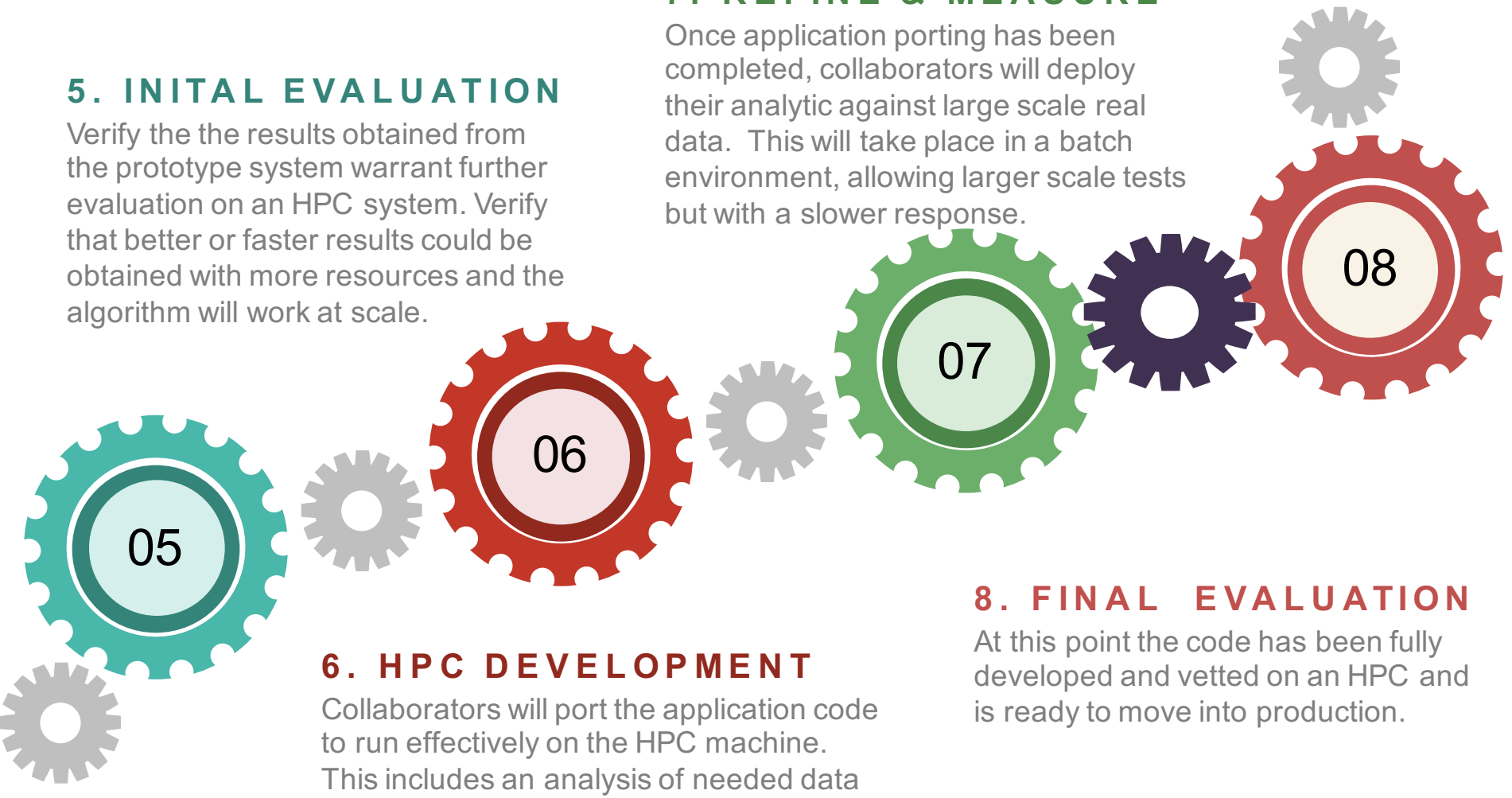
HPC Workflow

5. INITAL EVALUATION

Verify the the results obtained from the prototype system warrant further evaluation on an HPC system. Verify that better or faster results could be obtained with more resources and the algorithm will work at scale.

7. REFINE & MEASURE

Once application porting has been completed, collaborators will deploy their analytic against large scale real data. This will take place in a batch environment, allowing larger scale tests but with a slower response.



6. HPC DEVELOPMENT

Collaborators will port the application code to run effectively on the HPC machine. This includes an analysis of needed data and working with data owners to ensure data availability on the HPC.

8. FINAL EVALUATION

At this point the code has been fully developed and vetted on an HPC and is ready to move into production.

Overview of Operational Use Cases

VARR

- Vulnerability Awareness and Recommended Risk Remediation
- Provide **risk estimates** and **recommend courses of action** that maximize risk reduction using **modeling and simulation** techniques

MADHAT

- Multi-dimensional Anomaly Detection fusing HPC, Analytics and Tensors
- Identify **malicious network behavior** using **tensor decompositions** optimized for HPC environments with **deep learning** techniques

GALILEO

- Generalized Low-Entropy Mixture Model
- Enable **streaming anomaly detection** and near-neighbor detection of known malicious behavior using **clustering** techniques

Summary

- **Reduce barriers to data and computing resources**
- **Beginning of the effort**
  - Initial research results expected in June 2018
  - Seeking transition partners
- **Engaging broader community**
  - Novel research ideas
  - Development of benchmarks

Questions?

Leslie C. Leonard, PhD
Cybersecurity Research Lead
Leslie.Leonard@hpc.mil

William J. Glodek
william.j.glodek.ctr@mail.mil

Abbreviations and Acronyms

TERM	DEFINITION
API	Application Program Interface
DoD	Department of Defense
DREN	Defense Research and Engineering Network
ERDC	Engineer Research and Development Center
ESG	Executive Steering Group
HACSAW	HPC Architecture for Cyber Situational Awareness
HPC	High Performance Computing
HPCMP	High Performance Computing Modernization Program
HTTP	Hyper Text Transfer Protocol
IAP	Internet Access Point
MET	Mission Essential Task
SA	Situational Awareness
SDP	Service Delivery Point