

Cleared for Public Release

U.S. Department of Homeland Security | Science and Technology Directorate

Homeland Security Site Update: From Science to Operations

Briefing to HPC User Forum/HPC Site Updates



Science and
Technology

Amy E. Henninger, PhD, Senior Advisor and Branch Lead for Advanced Computing
Brian Henz, PhD, Senior Advisor for AI and Director of AI Technology Center
Technology Centers Division
Office of Science and Engineering
Science and Technology

Thursday, 7 September, 2023

Briefing Objectives and Outline

OBJECTIVES

Provide HPC User Forum attendees insight into:

- S&T's traditional HPC work.
- S&T's future requirements for HPC work.

OUTLINE

- Slide 3: GDNA Cluster
- Slide 4: Neutron Cluster
- Slide 5: Foundation Models
- Slide 6: AI Workflows
- Slide 7: Summary and Conclusions.

S&T's work in HPC space has traditionally been limited to classic computationally-intensive scientific applications.

GDNA Cluster (Genomic Data, Network and Analysis) National Bioforensic Analysis Center (NBFAC) National Biodefense Analysis & CMs Center (NBACC)



GDNA High-Performance Compute Cluster

- Highlights
 - More than 7500 CPU threads across 35 servers
 - 65 Terabytes of RAM
 - 1 Petabyte primarily SSD local disk
 - 3 Petabyte Panasas storage system
 - Running CentOS 7 and Ubuntu 20.04

Compute Nodes

22 AMD EPYC nodes, each with 256 CPU Threads, 1 TB RAM, 25 TB SSD storage

Large Memory Nodes

7 Intel Xeon nodes, each with 224 CPU threads, 3-6 TB RAM, 26 TB SSD storage

Custom Nodes

Application, Test, Backup, Head, and GPU nodes with 14,000+ GPU cores

Network

Cisco 10-40 GbE fiber infrastructure
Dual-10G bonded connections for each node
Air-gapped from Internet



UNCLASSIFIED

Storage

3 Petabyte Panasas ActiveStor Ultra storage system combining HDD, SSD, and NVDIMM storage

- Used primarily for the storage, assembly, and analysis of DNA sequencing data.
- Metagenomic Analysis
 - Identification of all biological contents in a forensic sample.
 - Analysis of >1M generated sequence reads
 - Build large, custom, comprehensive sequence DBs on GDNA to analyze sequences



Neutron Cluster @ National Urban Security Technology Laboratory (NUSTL)

- Runs the Los Alamos radiation transport code MCNP(6.2).
<https://mcnp.lanl.gov/>.
- Originally used to calculate the flux and energy distribution of subatomic particles, especially neutrons, as a way of detecting hidden nuclear weapons and special nuclear material.
- More recently, used to screen populations for radioactive contamination after a radioactive material release such as a dirty bomb.



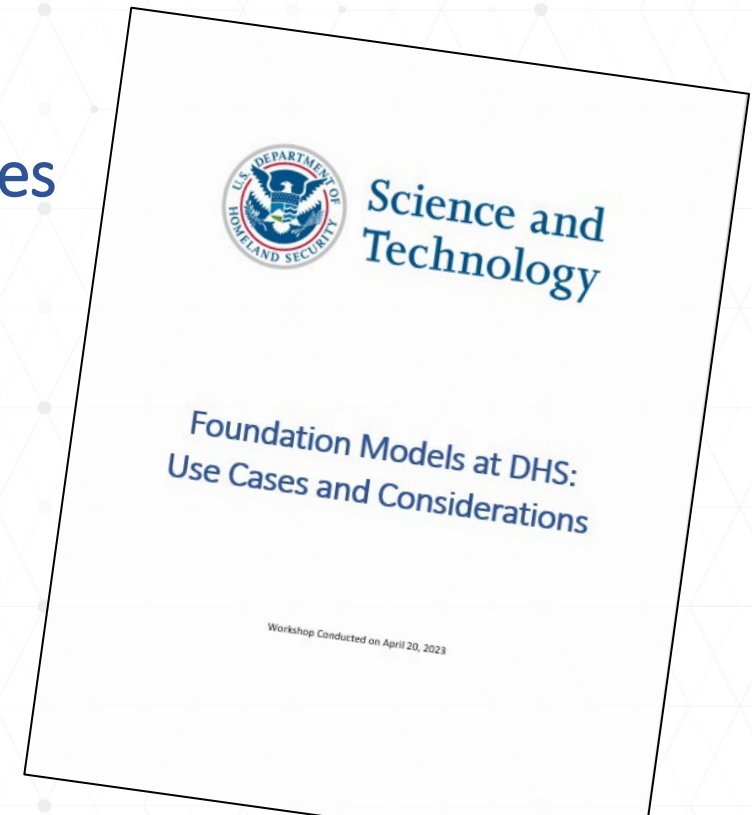
- Implementing CBRNE countermeasures
- Conducting risk assessment of cargo



Neutron Cluster is about to be retired and decommissioned.

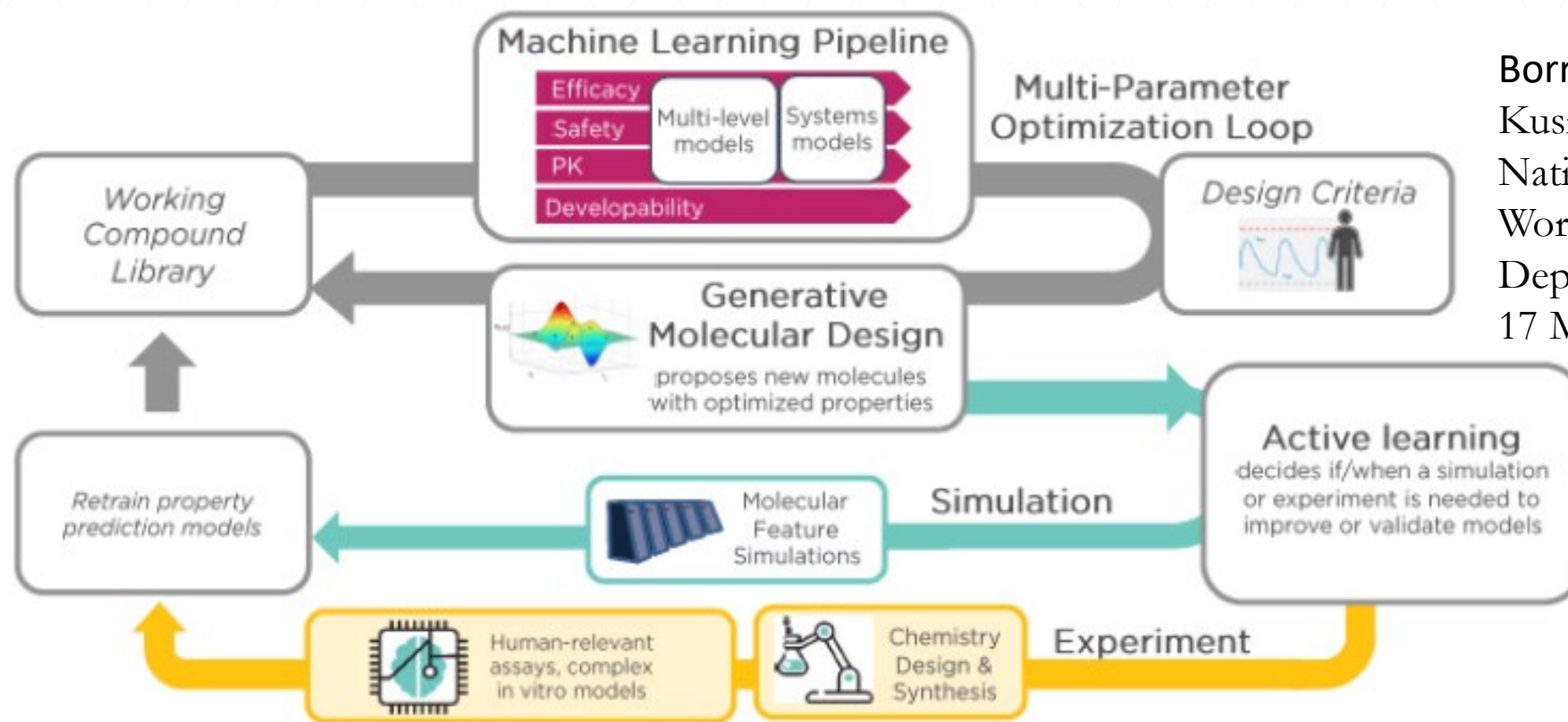
Foundation Models @ DHS: Potential Use Cases

- Law Enforcement to Include Digital Forensics and 911 Services
- Smuggling, Trafficking, Exploitation, and Illegal Activities at the Border
- Non-Intrusive Inspection and Scanning
- Emergency Management
- Genomics and Drug Discovery
- Immigration Services
- Cybersecurity
- Biometrics
- Business Applications



“By collapsing data and technology across use cases, foundation models benefit from increases in the scale and scope of datasets to become more capable and from economies of scale in workflow to become more efficient.”

AI Workflows and Integration with Simulations



Borrowed from:
Kusnezov, D. (2020).
National Academies
Workflow Discussion.
Department of Energy.
17 March, 2020.

“In the years to come, ideally, workflows will just be how we do day to day business. Data will drop into curated and labeled environments and naturally fall into the update cycles of learning, driving models, simulation and experiments in equal ways. AI will become the fourth pillar, augmenting experiment, theory and simulation so that all are part of the fluid tools we apply for solving problems important to us.”

D. Kusnezov, 17 March, 2020.

Summary and Conclusions

- While DHS has traditionally been a modest user of HPC iso S&T analytic applications, advances in machine learning and vast data opportunities impel DHS to look at using HPC on new AI workloads.
- Foundation models have demonstrated data and intelligence fusion that is a primary mission of DHS.
- As we establish foundation models and other AI-based capabilities to support the DHS mission, we maintain a focus on shared data processing requirements and simulations.
- DHS continues to watch emerging computing paradigms such as quantum computing, large scale conventional-quantum computing hybrids and their potential impacts on adversarial use of AI.

Questions?