SCIENCE USE CASES

Katherine Holcomb, UVA

We have received enthusiastic responses from the application research community looking forward to using this cyberinstrument. Illustrative science use cases were collected from the Virginia participating institutions. These are by no means exhaustive, but provide a sample of the many and varied research projects that will benefit from this unique computing environment.

The science use cases span multiple axes of potential applications using the proposed cyberinstrument, ranging from the security requirement axis, the performance demand axis, and diverse domain areas.

On the security axis, use cases span the range of protection from completely non-secure, to secure executable but non-secure data, through HIPAA and FISMA to network-permitted ITAR projects. Many of the non-secure projects also would benefit from data caching, where we would mirror a data repository. Each of these examples requires one or more petabytes of storage and would demand fast access from the computing hardware. An example is **UVA Department of Environmental Sciences**, where Prof. Kevin Grise works with state-of-the-art global climate models. Data from standard runs are freely provided to the broader climate science community for research and analysis on the Earth System Grid Federation (ESGF, URL). Large research institutions around the world are starting to create local mirrors of ESGF data, to allow their researchers easier and faster access to large quantities of data. Another example is **UVA: Center for Public Health Genomics.** The Center is involved in several "omics" projects, including **Multi-Ethnic Study of Atherosclerosis (MESA)**, which is a longitudinal study of subclinical cardiovascular disease and risk factors that predict progression to clinically overt cardiovascular disease or subclinical disease. This is a nationwide effort to perform multiple 'omics technologies on selected samples from various studies under the NHLBI portfolio. The current approach being employed by most sites is to provide information on every base in the genome, with significant expansion of size and storage requirements.

Sensitive data generally tends to make lesser demands on storage and access speed but is still of importance to management of research projects. These data include forensic data, in the project **Statistical Analysis of Forensics Data** at **UVA**, which is a growing focus of Prof. Karen Kafadar's research. Her group expects to encounter data sets from crime labs that require data protection. The proposed system would allow her team to carry out their research on both protected and unprotected data in a seamless manner. A project at **George Mason University** on **Using Scientific Research to Shape Crime and Justice Policies** is being implemented through the Center for Evidence Based Crime. The project seeks to make scientific research a key component in decisions about crime and justice policies.

Many projects also span the data requirements, such as **Virginia Tech's Mass Spectrometry Incubator (VT-MSI).** The VT-MSI provides mass spectrometry data to approximately 40 research groups on campus, analyzing thousands of samples per year. Datasets provided to the investigators range in size from a few megabytes to several terabytes. As a multiuser facility, we see multiple levels of access requirements as well as multiple needs for data security covering the gamut from private (PI is not ready to share publicly) to ITAR.

We also have a number of projects depending directly upon sensitive patient data, as well as simulation for health-provider training that would benefit from using actual patient data. for example, **Old Dominion University's Training for Military Health Practitioners** project. The PI, Dr. Andrea Parodi, is researching military trauma, burns, and patient safety using TeamSTEPPS, a form of crew resource management.

Additional work focuses on the use of simulation in medical education, serious games for clinicians, and the uses of computational modeling and simulation in healthcare settings.

Along the range of performance requirements. Several sensitive-data projects require little computing power. For example, the project **Contentious Interpersonal Interactions** at the **College of William and Mary** collects a large amount of confidential data related to subjects' personality, psychology, physiology, and political beliefs for the study of contentious interactions.

Others require significant computational work. For example, **Virginia Tech: In-situ and Remote Sensing of the Ocean**. Eric Paterson, Department of Aerospace and Ocean Engineering. Information about the ocean is important for weather and climate forecasting, routing of ship traffic, monitoring of pollution, evaluating the health of fisheries, and national defense. Sensors of numerous types can be fixed to structures and buoys, or flown on spacecraft, aircraft, ships, and submersibles In-situ sensing is equally complex, and includes numerous sensing modalities. Large-scale computational fluid dynamics (CFD) and computational electromagnetics (CEM) is used to perform physics-based simulations of ship wakes in the ocean.

The subject matter of our use cases covers nearly all areas of science and engineering. Here, we provide examples of projects involving human subjects. This includes several projects using patient or other health data, but other types of Personally Identifiable Information are also needed, such as psychological profiles, criminal histories, addresses, and so forth. We also have cases utilizing proprietary corporate data including **George Mason University's** project titled **Protecting Proprietary, Export Controlled, and Covered Defense Information in a Cross-institutional DARPA-Sponsored Collaboration.** GMU's Center for Assurance Research & Engineering (CARE) conducts vital research on the most critical cybersecurity issues in our world today, transforming research into innovative solutions that increase security in real-world settings. Center researchers recently began work under approximately $4 million in grants from the Defense Advanced Research projects Agency (DARPA). This work, performed in collaboration with public and private universities, including Columbia University and Penn State University and with private companies, will develop a "moving target" defense against distributed denial-of-service attacks.

The diversity of our science cases demonstrates the tremendous potential of our proposed instrument. Projects now at best awkward and at worst nearly impossible would become far more productive. The resulting benefit to science and engineering throughout the Commonwealth of Virginia and the world would be enormous.