Application Title: A High-Performance Data Analysis System to Enable Transformation of Big Data and Computational Research and Education at Texas A&M

Lead contact for RDF Application: Honggao Liu, Director, **High Performance Research Computing (HPRC)**, honggao@tamu.edu, 979-845-2561

Key Participating Units:
- **College of Engineering/TEES:** Aerospace, Biomedical, Chemical, Computer Science, Civil, Electrical, Engineering Technology & Industrial Distribution, Mechanical, Materials Science and Engineering, Petroleum;
- **College of Science**: Chemistry, Physics & Astronomy, Statistics, Mathematics;
- **College of Geosciences**: Oceanography, Atmospheric Sciences, Geography;
- **Health Science Center:** Health Policy & Management; Epidemiology & Biostatistics
- **College of Veterinary Medicine & Biomedical Sciences:** Veterinary Physiology and Pharmacology, Veterinary Integrative Biosciences;
- **College of Agriculture & Life Sciences**: Biochemistry and Biophysics;
- **College Mays Business School:** Marketing;
- **Texas A&M Transportation Institute:** Infrastructure Protection;
- **Texas A&M University at Galveston:** Marine Biology;
- **Centers**: AgriLife Genomics and Bioinformatics Service, Laboratory for Molecular Simulation.

Anticipated Request Amount ($): 1,800,000

Executive summary of the intended application to utilize Research Development Funds:

This proposal seeks funding for acquisition of a new research instrumentation designed as a high-performance data analysis and computing system for both big data analytics and computational sciences. The proposed system will be equipped with the latest technologies such Intel Xeon Skylake processors, NVIDIA Volta GPUs, SSD and NVMe storage, and interconnected with Intel Omni-Path Architecture (OPA) fabric. This system will provide shared resources for all academic disciplines with the goal of Texas A&M System-wide and national impact while building next generation data science and research computing workforce, and participating regional and national cyberinfrastructure activities.

The proposed data analysis and computing system with innovative data processing capabilities and huge high-performance data storage, will greatly enhance the research strength and competitiveness of Texas A&M in computational and data-enabled science and engineering, and leverage the university's efforts and resources for the recently created **Texas A&M Institute of Data Science**. It will serve as a critical research environment to advance projects dependent on big data analytics and high performance computing, and directly support more than 30 investigators from 25 different departments involved in this project and their groups. It will be the focal resource for more than 10 workshops and over 30 courses per year intent on training a new generation of computational and data science researchers to use the system to enable new discovery. The system will allow Texas A&M to take the next step in data science and data-intensive research, and produce a workforce of researchers ready for a new generation of national leadership supercomputers.

The payoff from this project is significant, and it has a wide range of broad impacts, from transforming research and education, advancing workforce development to disseminating research results. The HPRC has been a dedicated resource for research and discovery at Texas A&M, and currently supports more than 1,800 users, including more than 450 faculty members. HPRC promotes emerging technology to researchers and assists them in using it for research and discovery. Through support for "long tail" science, and availability of the new capacity to faculty, staff and students within the TAMU System, and training in its use as a tool for both research and education, the proposed system will foster much broader participation in STEM learning and research by traditionally underserved groups.