CIF21 DIBBs: Middleware and High Performance Analytics Libraries for Scalable Data Science

PI: G. Fox, Co-PIs: Madhav Marathe, Shantenu Jha, Judy Qiu, Fusheng Wang
Institutions: Arizona State, Indiana (lead), Kansas, Rutgers, Stony Brook, Virginia, and Utah.

Award #: **1443054**

- **Community Driven High Performance Big Data for bio-physical applications based on HPC, distributed systems, network science, GIS and machine/deep learning**
- **Pilot Jobs, Twister2, Harp-DAAL HPC-Cloud convergent middeware (HPCforML)** replaces Spark+Hadoop with ML and DL for streaming and batch applications
- **Polar Science:** Operational ML/DL to locate ice sheet boundaries and snow layers from radar data
- **Network Science:** parallel subgraph, anomaly and community detection algorithms for CINET
- **Epidemic analysis:** MLforHPC enhanced simulations TDEFSI for Influenza dynamics
  - Supporting the Coronavirus outbreak mitigation with simulation of different spreading scenarios and possible interventions
- **Public Health GIS:** Spatial big data query for opioid epidemic prevention and intervention
  - **Pathology:** DL based image analysis tools for level-set based image segmentation, 3D registration, reconstruction, and spatial analysis
- **Biomolecular Simulations:** PMDA parallelizes widely used MDAnalysis Python package for MD trajectory analysis
  - **MLforHPC** improves molecular dynamics simulation performance by factor $10^{4\text{-}5}$ for short time (recurrent) and long time (fully connected networks); roadmap for other applications