National Institute of Standards & Technology

Poster Session

Report

Tuesday, March 4th 2014
Common Themes

- Infrastructure
  - Open source
  - Proprietary

- Analytics
  - Heterogeneous data integration
  - Scaling existing analytics tools to big data

- Repositories & Storage
  - Earth Sciences
  - Biomedical
  - Imaging
Infrastructure

- Open source
  - Open Science Data Cloud (OSDC)
    - Allocations start at 16 cores and 1TB storage
  - MongoDB
    - NoSQL database solution
  - Research Data Alliance (RDA)
    - Bring together expertise across disciplines

- Proprietary
  - Teradata
    - Full stack solution (hardware + software)

- General Research
  - Managing web scale knowledge
    - Infrastructure for organizing and connecting computing resources and tools
  - Network traffic
    - Identify inefficiencies to optimize processing
Analytics

• Heterogeneous data integration
  – Health data harmonization
    • Many data types for each patient
    • Identify combination predictive of outcome
  – National Ecological Observatory Network (NEON)
    • Combining imaging data, sensor & other data for prediction

• Scaling existing analytics tools to big data
  – Image processing
    • Segmentation & image stitching
  – Popular open source analytics tools & real time computation
    • Scaling R, stata, others to big data using open source (e.g. Cloudera Hadoop) and proprietary (e.g. Revolution, Teradata, etc) tools
  – Experimental design
    • Space filling design for big data
Repositories and Storage

- NEON
  - Open source environmental scale observatory

- Cell Imaging
  - Collection and storage TB scale temporal images

- Next Generation Sequencing (NGS)
  - Utilization of AWS to make NGS data publicly available
  - Genome in a Bottle Consortium

- Teradata
  - Hardware infrastructure for distributed data storage and computing