**NBD(NIST Big Data) Requirements WG Use Case Template Aug 11 2013**

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | Radar Data Analysis for CReSIS | |
| **Vertical (area)** | | Scientific Research: Polar Science and Remote Sensing of Ice Sheets | |
| **Author/Company/Email** | | Geoffrey Fox, Indiana University gcf@indiana.edu | |
| **Actors/Stakeholders and their roles and responsibilities** | | Research funded by NSF and NASA with relevance to near and long term climate change. Engineers designing novel radar with “field expeditions” for 1-2 months to remote sites. Results used by scientists building models and theories involving Ice Sheets | |
| **Goals** | | Determine the depths of glaciers and snow layers to be fed into higher level scientific analyses | |
| **Use Case Description** | | Build radar; build UAV or use piloted aircraft; overfly remote sites (Arctic, Antarctic, Himalayas). Check in field that experiments configured correctly with detailed analysis later. Transport data by air-shipping disk as poor Internet connection. Use image processing to find ice/snow sheet depths. Use depths in scientific discovery of melting ice caps etc. | |
| **Current**  **Solutions** | **Compute(System)** | | Field is a low power cluster of rugged laptops plus classic 2-4 CPU servers with ~40 TB removable disk array. Off line is about 2500 cores |
| **Storage** | | Removable disk in field. (Disks suffer in field so 2 copies made) Lustre or equivalent for offline |
| **Networking** | | Terrible Internet linking field sites to continental USA. |
| **Software** | | Radar signal processing in Matlab. Image analysis is MapReduce or MPI plus C/Java. User Interface is a Geographical Information System |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | Aircraft flying over ice sheets in carefully planned paths with data downloaded to disks. |
| **Volume (size)** | | ~0.5 Petabytes per year raw data |
| **Velocity**  **(e.g. real time)** | | All data gathered in real time but analyzed incrementally and stored with a GIS interface |
| **Variety**  **(multiple datasets, mashup)** | | Lots of different datasets – each needing custom signal processing but all similar in structure. This data needs to be used with wide variety of other polar data. |
| **Variability (rate of change)** | | Data accumulated in ~100 TB chunks for each expedition |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues)** | | Essential to monitor field data and correct instrumental problems. Implies must analyze fully portion of data in field |
| **Visualization** | | Rich user interface for layers and glacier simulations |
| **Data Quality** | | Main engineering issue is to ensure instrument gives quality data |
| **Data Types** | | Radar Images |
| **Data Analytics** | | Sophisticated signal processing; novel new image processing to find layers (can be 100’s one per year) |
| **Big Data Specific Challenges (Gaps)** | | Data volumes increasing. Shipping disks clumsy but no other obvious solution. Image processing algorithms still very active research | |
| **Big Data Specific Challenges in Mobility** | | Smart phone interfaces not essential but LOW power technology essential in field | |
| **Security & Privacy**  **Requirements** | | Himalaya studies fraught with political issues and require UAV. Data itself open after initial study | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | | Loosely coupled clusters for signal processing. Must support Matlab. | |
| **More Information (URLs)** | | http://polargrid.org/polargrid  https://www.cresis.ku.edu/  See movie at http://polargrid.org/polargrid/gallery | |
| **Note:** <additional comments> | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Use Case Stages** | **Data Sources** | **Data Usage** | **Transformations  (Data Analytics)** | **Infrastructure** | **Security & Privacy** |
| **Radar Data Analysis for CReSIS (Scientific Research: Polar Science and Remote Sensing of Ice Sheets)** | | | | | |
| **Raw Data: Field Trip** | Raw Data from Radar instrument on Plane/Vehicle | Capture Data on Disks for L1B.  Check Data to monitor instruments. | Robust Data Copying Utilities.  Version of Full Analysis to check data. | Rugged Laptops with small server (~2 CPU with ~40TB removable disk system) | N/A |
| **Information:**  **Offline Analysis L1B** | Transported Disks copied to (LUSTRE) File System | Produce processed data as radar images | Matlab Analysis code running in parallel and independently on each data sample | ~2500 cores running standard cluster tools | N/A except results checked before release on CReSIS web site |
| **Information:**  **L2/L3 Geolocation & Layer Finding** | Radar Images from L1B | Input to Science as database with GIS frontend | GIS and Metadata Tools  Environment to support automatic and/or manual layer determination | GIS (Geographical Information System).  Cluster for Image Processing. | As above |
| **Knowledge, Wisdom, Discovery:**  **Science** | GIS interface to L2/L3 data | Polar Science Research integrating multiple data sources e.g. for Climate change.  Glacier bed data used in simulations of glacier flow |  | Exploration on a cloud style GIS supporting access to data.  Simulation is 3D partial differential equation solver on large cluster. | Varies according to science use. Typically results open after research complete. |