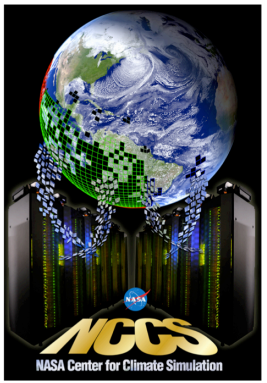
ABoVE Science Cloud
Support for the Pre-ABoVE Projects

NASA Center for Climate Simulation (NCCS) Code 606.2
Carbon Cycle & Ecosystems Office
Mark Carroll (Code 618)
Tatiana Loboda (University of Maryland)



Goal and Concept

Goal

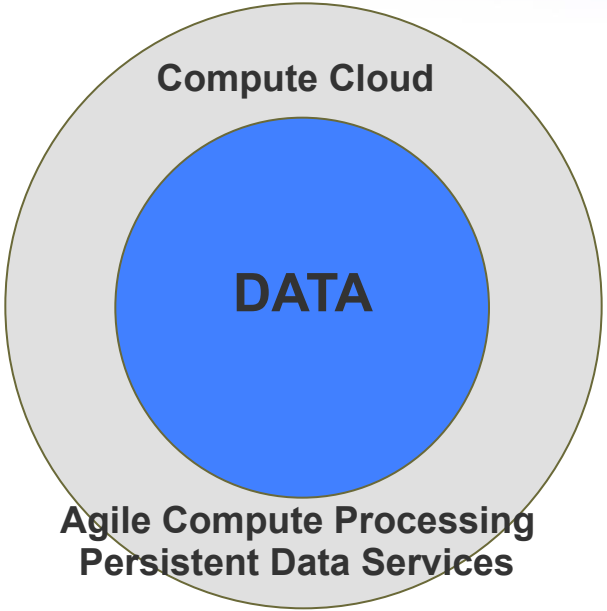
- Demonstrate support for the ABoVE field campaign and the Carbon Cycle and Ecosystems Office by providing direct support for the five pre-ABoVE projects http://above.nasa.gov/

Concept

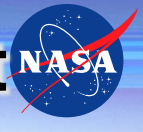
- Unified Data Analysis Platform that provides a co-location of data, compute, data management, and data services
- Provided by the NASA Center for Climate Simulation (Code 606.2) in Building 28 at Goddard Space Flight Center

Data storage surrounded by a compute cloud

- Large amount of data storage, high performance compute capabilities, very high speed interconnects



Proposed Capabilities for Pre-ABoVE Support



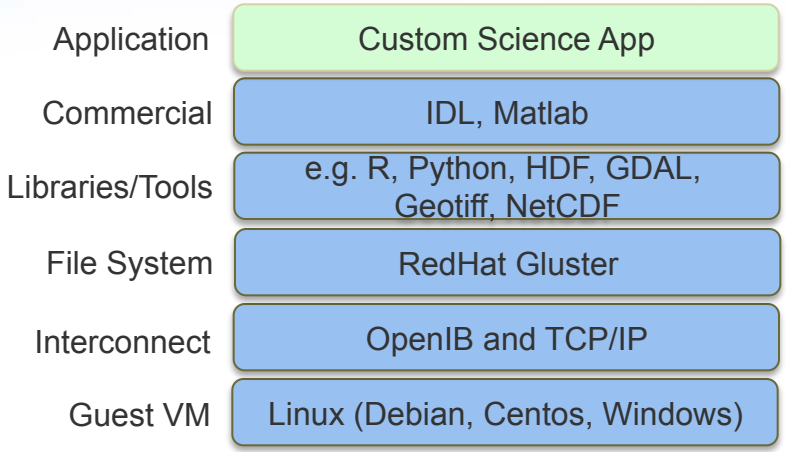
Storage

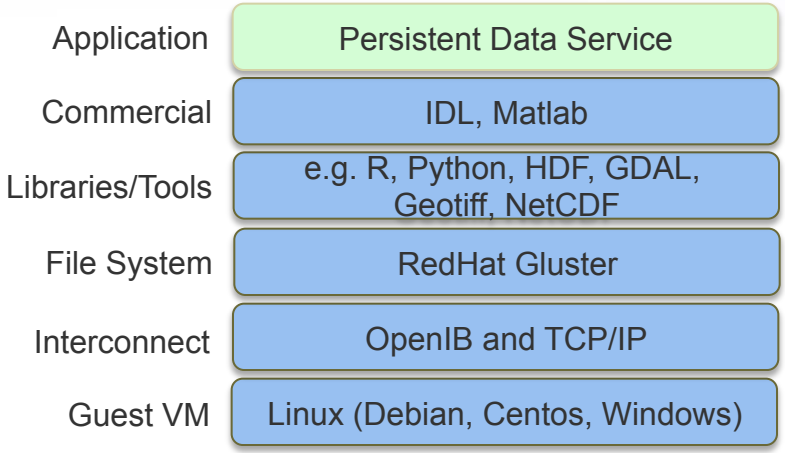
- **Start by using existing data service storage servers (possibly augment)**
- **Petabytes of disk storage**

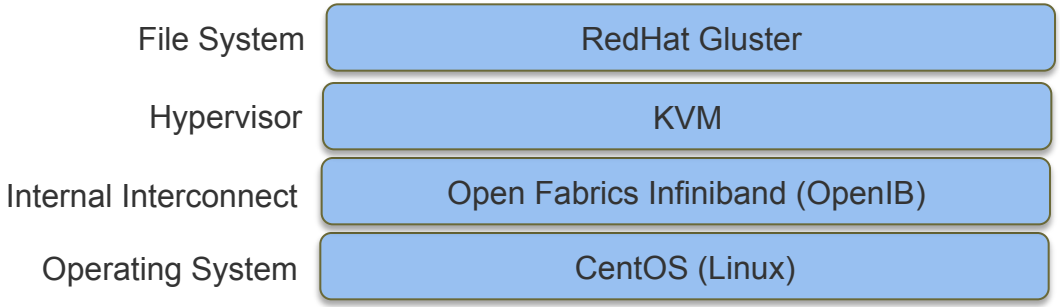
Compute

- **Make use of existing servers that are being upgraded**
  - 11 Dell C6100 Nodes
  - Dual-socket, hex-core 2.8 GHz Intel Xeon Westmere processors
  - 11x12 = 121 total cores
  - 24 GB of RAM per node
- **Infiniband interconnect for internal network**
- **1 GbE for external network (to start with)**
- **System is architected to be easily expanded for future requirements**
- **Housed in building 28 data center**
- **Falls under existing NCCS security plan**

Example Software Stacks and Responsibilities







Responsibilities

Scientists

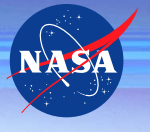
NCCS

Simple Concept Of Operations

Principal Investigators and Users will

- **Identify data sets to be copied into the science cloud and get support for data management, metadata, naming conventions, etc.**
- **Have use of private “virtual machine” work environments to run their own code and generate products for their own use and evaluation**
- **Bring their algorithms to the compute cloud**
- **Manage those algorithms in the cloud**
- **Manage the outputs and results of the algorithms**
- **Receive assistance for the creation and management of persistent data services when data products are ready to be shared within the team and to outside partners**

Simple Concept Of Operations



Science Cloud will

- **Support account creation**
- **Manage private virtual machines for the scientists**
- **Support the movement of data into the cloud**
- **Provide support for the scientists on accessing and processing in the science cloud**
- **Support the creation and management of persistent data services within the compute cloud**
- **Assist users in trouble shooting in the event of errors with processing data**
- **Provide guidance for the creation, search, discover, and use of metadata**

Science Benefit of a Science Cloud

Carroll et al Landsat water maps for 3 epochs:

- Expanded storage availability allowing denser stacks of data (expanding temporally into shoulder seasons)
- Data processing times cut down from 6 – 9 months to 6 – 10 weeks
- Extra time to be used for expansion of QA/verification of products
- Expansion of the region from the “tier 1 primary region” into tier’s 2 and 3 as outlined in proposal, including the expanded High Arctic region defined by the SDT
- Enables possible expansion of the time domain, in the future, from epochs to full time series of all available data

Science Benefit of a Science Cloud

Loboda et al MODIS cloud dynamics and burned area mapping

- Dramatically reduces download time by utilizing multiple high speed connections simultaneously
- Data processing times will be reduced by
  - No need to cycle through inputs, all can stay online at once
  - Multiple instances can be run simultaneously
- Enables the option for reprocessing as the algorithm evolves
- Extra time for QA/verification of the products
- Accelerates completion of circumpolar data product
- Solves data storage and data management issues freeing up personnel time
- Reduction of effort because Loboda will also use the Landsat from Carroll

Science Benefit of a Science Cloud

- Walker et al Alaska Vegetation Atlas is being carefully created with assistance from GINA (Geographic Information Atlas of Alaska). It would be a test-case for "Data as a Service" linkages between GINA and the ABoVE Science Cloud
- ReSalt and CO2 data assimilation framework: Investigators have been contacted about formatting and metadata creation for sharing of their products upon completion
- Enhance CCE Office use of ESRI ArcGIS and exploration of RAMADDA
  - Desktop on super fast hardware with essentially unlimited disk storage
  - Online (in ESRI's cloud) becomes Portal (in ABoVE Science Cloud)
  - Data as a Service via ArcGIS Server

Discussion of Implications

ABoVE Concise Experiment Plan

Remote sensing activities during ABoVE will include:

1. Use of existing information products from satellite and airborne RS data;
2. Use of proven methodologies to create additional information products from historical archives of satellite and airborne RS data; and
3. Use field observations to develop, calibrate, and/or validate new information products from existing and new airborne and spaceborne and airborne remote sensors.

ABoVE NRA in ROSES 14

Assessment of performance and long term availability by June 2014