High Resolution Nature Runs and the Big Data Challenge

iCAS 2015
Annecy, FR
17 Sept, 2015

W. Phillip Webster, John Schnase & Daniel Duffy
NASA Goddard Space Flight Center (GSFC)
NASA Center for Climate Simulation (NCCS)

Provides an integrated high-end computing environment designed to support the specialized requirements of Climate and Weather modeling.

- High-performance computing, data storage, and networking technologies
- High-speed access to petabytes of Earth Science model and observation data
- Collaborative data sharing and publication services
- Advanced Data Analytics Platform (ADAPT) – High Performance Science Cloud

Primary Customers (NASA Climate Science)

- Global Modeling and Assimilation Office (GMAO)
- Goddard Institute for Space Studies (GISS)

High-Performance Science

- [http://www.nccs.nasa.gov](http://www.nccs.nasa.gov)
Global Climate Modeling

Takes in small input and creates large output

• Using relatively small amount of observation data, models are run to generate forecasts
• Fortran, Message Passing Interface (MPI), large shared parallel file systems
• Rigid environment – users adhere to the HPC systems

Example: GEOS-5 Nature Run (GMAO)

• 2-year Nature Run at 7 KM resolution
• 3-month Nature Run at 3 KM resolution
• Generated about 4 PB of data (compressed)
• To be used for Observing System Simulation Experiments (OSSE’s)
• All data to be publically accessible
• ftp://G5NR@dataportal.nccs.nasa.gov/

Obs Data

Model
(Many 100K lines of code)

The Goddard Chemistry Aerosol Radiation and Transport (GOCART) model,
Courtesy of Dr. Bill Putman, Global Modeling and Assimilation Office (GMAO),
NASA Goddard Space Flight Center.
Typical Analysis Applications

Takes in large amounts of input and creates a small amount of output

- Using large amounts of distributed observation and model data to generate science
- Python, IDL, Matlab
- Agile environment – users like to run in their own environments
- Lots of data reads!

Examples

- Evaporative transport
  - Requires monthly reanalysis data sets for four different spatial extents
  - Decadal water predictions for the high northern latitudes for the past three decades
  - Requires 100,000+ Landsat images and about 20 TB of storage

Notional Architecture for Climate HPC

NCCS Local Area Network
10 GbE and 40 GbE

Mass Storage
• Tape ~45 PB
• Disk ~4 PB

Advanced Data Analytics Platform (ADAPT)
• Web services
• Designed for large scale data analytics
• Science Cloud
• HPC Technologies

High Performance Computing Cluster Discover
• Large Scale Models
• ~3.5 PF Peak
• >80,000 Cores

GPFS Shared File System
~33PB

• 1,000’s of cores
• Petabytes of storage
• Using decommissioned HPC systems
7-km GEOS-5 Nature Run
Overview

Nature Run Details

- 2-years: June 2005 – June 2007
- 7-km Global Resolution
- Non-Hydrostatic Cubed-Sphere Finite-Volume Dynamics
- RAS convection (limited by stochastic Tokioka scheme)
- Resolve meso-scale weather and convective clusters
- High-resolution constituent transport
- Executing at NASA GSFC/NCCS: throughput of 11-days/day

Surface Boundary Conditions & Emissions

- 0.25-degree combined Reynolds/OSTIA sst and sea-ice
- Fossil fuel CO/CO2 emissions from 0.1-degree EDGAR inventory
- Land CO2 fluxes from CASA-GFED at 0.1-deg with MODIS EVI
- QFED daily varying biomass burning emissions
- AEROCOM volcanic SO2 emissions and injection heights
- GOCART includes mixing, chemistry, and deposition of key aerosol types including sulfates, dust, and black carbon
- GOCART aerosols are radiatively coupled with the dynamics

Surface CO2 Concentration:
The surface concentration of CO2 highlights the fidelity of local emission centers and constituents being dispersed within the regional meso-scale flow.

Column CO2 Concentration:
Surface emissions are lifted throughout the column and transported beyond the regional scales within the planetary flow.
## Nature Run Statistics

<table>
<thead>
<tr>
<th></th>
<th>7.0 km</th>
<th>3.5 km</th>
<th>1.75 km&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resolution</strong></td>
<td>7.0 km</td>
<td>3.5 km</td>
<td>1.75 km</td>
</tr>
<tr>
<td><strong>Surface Grid (10^6)</strong></td>
<td>12.24</td>
<td>48.98</td>
<td>195.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>x 72 levels</strong></td>
<td>48.98</td>
<td>195.91</td>
<td></td>
</tr>
<tr>
<td><strong>Simulation Time</strong></td>
<td>2 years</td>
<td>2 months&lt;sup&gt;2&lt;/sup&gt;</td>
<td>7 days</td>
</tr>
<tr>
<td><strong>Cores</strong></td>
<td>7,200</td>
<td>21,000</td>
<td>28,000</td>
</tr>
<tr>
<td><strong>Days</strong></td>
<td>61 days</td>
<td>12 days</td>
<td>7 days</td>
</tr>
<tr>
<td><strong>Core hours (10^6)</strong></td>
<td>10.5</td>
<td>6</td>
<td>4.7</td>
</tr>
<tr>
<td><strong>Output (PB) compressed</strong></td>
<td>3</td>
<td>0.5&lt;sup&gt;3&lt;/sup&gt;</td>
<td>0.25&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>1</sup>Demonstration only
<sup>2</sup>Planning 1 year run this fall
<sup>3</sup>4 PB expected from 1 year run
<sup>4</sup>Selected fields only
While the nature run does not produce specific weather events for the period of 2005-2007, it should produce realistic synoptic events consistent with the types of events observed over that period. We have examined this simulation for extreme events typical of the nature run simulation and present a preview of these events below.

**Selected events found in the Nature Run**

November 4-9, 2005 atmospheric river

Global tropical cyclones (2005 and 2006)

Selected tropical cyclone events

May 2006 US Midwest Mesoscale Convective Complexes

January 13-21, 2007 Eastern US winter storms

Northern Hemisphere January 2006/2007 storm tracks

Global carbon concentration (2006 annual cycle)

Sierra Negra volcanic eruption in the Galapagos (October 22-30, 2005)
7-km GEOS-5 Nature Run
November 4-9, 2005 Atmospheric River
7-km GEOS-5 Nature Run
November 4-9, 2005 Atmospheric River

7-km GEOS-5 NR Total Rainfall (November 4-9, 2005)

Percent of the total November-2005 Rainfall that fell during the November 4-9, 2005 atmospheric river event.
7-km GEOS-5 Nature Run
2006 Column Integrated CO2 (colors) and CO (white)
Highest Resolution Simulation by US Model

- 1.75 KM Global Simulation with GEOS-5; simulated 1 climate day per wall clock day
- This non-hydrostatic, high resolution simulation incorporated a new 2-moment microphysics scheme, interactive aerosols, and produced very realistic weather features; pushing toward explicit cloud resolving resolutions
# The Growth of Climate Data

<table>
<thead>
<tr>
<th>35-year Reanalysis</th>
<th>Resolution</th>
<th>Data Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>MERRA and MERRA2 – Current NASA reanalysis</td>
<td>50 KM</td>
<td>400 TB</td>
</tr>
<tr>
<td>Current NASA operational resolution</td>
<td>25 KM</td>
<td>1.6 PB</td>
</tr>
<tr>
<td>Current NOAA operational resolution; 15 of 35 years will be complete by this fall (2015 – THIS YEAR)</td>
<td>12 KM</td>
<td>6.4 PB</td>
</tr>
<tr>
<td>Cloud permitting models, still parameterized (currently have a 2 year simulation)</td>
<td>7 KM</td>
<td>26 PB</td>
</tr>
<tr>
<td>Current high resolution climate runs (currently have a 3 month simulation)</td>
<td>3 KM</td>
<td>102 PB</td>
</tr>
<tr>
<td>Resolving deep convection – currently simulate 1 model day per wall clock day (model climate in real time)</td>
<td>1 KM</td>
<td>410 PB</td>
</tr>
<tr>
<td>Cloud permitting – need cloud and coupled ocean atmosphere models</td>
<td>0.75 KM</td>
<td>1.6 EB</td>
</tr>
</tbody>
</table>
Typical Analysis Applications

Takes in large amounts of input and creates a small amount of output

• Using large amounts of distributed observation and model data to generate science
• Python, IDL, Matlab
• Agile environment – users like to run in their own environments
• Lots of data reads!

Examples

• Evaporative transport
  • Requires monthly reanalysis data sets for four different spatial extents
  • Decadal water predictions for the high northern latitudes for the past three decades
  • Requires 100,000+ Landsat images and about 20 TB of storage

Analysis
(100’s of lines of code)

Representative Landsat image, false color composite, from near Barrow, AK; Courtesy of Mark Carroll (618).

Yukon Delta Alaska; courtesy of Landsat
http://landsat.visibleearth.nasa.gov/view.php?id=72762
Web/Data services are no longer adequate

The primary method of obtaining data to answer any of the previous questions is to find the source of the data and download the data.

- Data movement and wrangling has been estimated to take as much as 60%-80% of a scientist’s time
- Given the sizes and complexities of the data sets, this has become intractable.
- People are limiting their science, limiting their questions, using older, smaller data sets.
- In some cases, people who need the data don’t know where to find it or how to use it.

So what do we do now?

- Build pervasive, federated analytical services …
Advanced Data Analytics Platform (ADAPT)  
“High Performance Science Cloud”

Adjunct to the NCCS HPC environment
• Lower barrier to entry for scientists
• Customized run-time environments
• Reusable HPC/Discover hardware

Expanded customer base
• Scientist brings their analysis to the data
• Extensible storage; build and expand as needed
• Persistent data services build in virtual machines
• Create purpose built VMs for specific science projects

Difference between a commodity cloud
• Platform-as-a-Service that comes close to matching HPC levels of performance
• Supported user environment – scientists don’t have to be IT or cloud experts
• Critical Node-to-node communication – high speed, low latency
• Shared, high performance file system
• Management and rapid provisioning of resources
Climate Analytics-as-a-Service (CAaaS) combines large-scale data management, high-performance, storage-side computing and a domain-specific application programming interfaces to deliver climate analytic capabilities to a broad range of applications and customers (not just climate experts).

Create a service that contributes to a global network of sector-specific data, driving innovation and discovery.
How to build CAaaS

Data

Relevance and Collocation

Data have to be significant, sufficiently complex, and physically or logically co-located to be interesting and useful …

Exposure

API

Convenient and Extensible and Generative

Capabilities need to be easy to use and facilitate community engagement and adaptive construction …

High-Performance Compute/Storage Fabric

Storage-proximal analytics with simple canonical operations

Data do not move, analyses need horsepower, and leverage requires something akin to an analytical assembly language …

Reference:

MERRA Analytics Service (MAS) – Instantiation of a CAaaS

**MERRA Reanalysis**

**Data**

*Relevance and Collocation*

Data have to be significant, sufficiently complex, and physically or logically co-located to be interesting and useful …

**Climate Data Services API**

**Convenient and Extensible**

Capabilities need to be easy to use and facilitate community engagement and adaptive construction …

**MERRA Analytic Services**

**High-Performance Compute/Storage Fabric**

*Storage-proximal analytics with simple canonical operations*

Data do not move, analyses need horsepower, and leverage requires something akin to an analytical assembly language …
Modern Era–Retrospective Analysis for Research and Applications

- Source: Global Modeling and Assimilation Office (GMAO)
- Input: 114 observation types (land, sea, air, space) into “frozen” numerical model. (~4 million observations/day)
- Output: a global temporally and spatially consistent synthesis of 26 key climate variables. (~418 under the hood.)
- Spatial resolution: 1/2° latitude × 2/3° longitude × 42 vertical levels extending through the stratosphere.
- Temporal resolution: 6-hours for three-dimensional, full spatial resolution, extending from 1979–Present.

~ 200 TB, but MERRA II is on the way…
High Performance Compute/Storage Fabric – Hadoop Cluster

**Hardware Configuration**
- 36 node Dell cluster (11.7 TF Peak)
- 576 total cores (Intel 2.6 GHz SandyBridge)
- 2,304 GB of RAM (64 GB per node)
- 1,296 TB of RAW storage (12 x 3 TB disk drives; 36 TB per node)
- FDR Infiniband (56 Gbps); TCP/IP speeds at 20 Gbps

**Software Configuration (Open Source)**
- Cloudera Hadoop v5.0+
- Centos base operating system
- Using virtual machines – test, pre-production, production all on the same hardware

**Storage Layout**
- Majority of storage allocated to production
- Each disk is its own separate volume within HDFS – helps to eliminate disk contention
We have created a small set of canonical near-storage, early-stage analytical operations that represent a common starting point in many analysis workflows in many domains. For example,

- Average, Maximum, Minimum, Sum, Count, Variance, Anomaly

Operations of the general form:

- \( \text{result} \leq \text{avg}(\text{var}, (t_0, t_1), ((x_0,y_0,z_0),(x_1,y_1,z_1))) \),

This call will return the average of a variable when given a variable name, temporal extent, and spatial extent ...

All without having to know the data formant (NetCDF) or even MapReduce!
Simple Arctic Boreal Related Example

QUESTION: Extract the average temperature by season for the year 1980 for the arctic region at every vertical height in the MERRA data (only surface temperatures shown below).
Surface Temperature Anomaly of the Independent Reanalysis Compared to the 34-Year Ensemble Average

- Work done by Denis Nadeau (NCCS)
- Temperature variability in the pictures is due to such things as topographical features, different cloud parameterizations within the models, etc.
- Departure of average reanalyses over summer 2010(JJA) shows that ECMWF surface temperature is generally colder than MERRA.
The Challenge of Hadoop

NASA science requirements
- Scientists are creating huge amounts of data, and they are not stopping!
- Still want to access those data sets through their typical way of analyzing data, i.e., a file system (POSIX)
- How do we use Hadoop or other object store environments?
  - Must create another copy of the data (sequence for Hadoop)
  - No longer a file system like interface (REST)
  - Yes, you can try to layer FUSE on top, but performance and reliability is sketchy (for now)
  - Even if you layer a file system on top of sequenced data, the scientists would not be able to read the data any longer

This is an intractable problem for us.
- What are we trying to do to solve this?
Using Native NetCDF with Hadoop

Working with George Mason University researchers (Phil Yang et al) to explore the potential of using Native NetCDF (non-sequenced data) in HDFS

- Zhenlong Li, Fei Hu, John Schnase, Daniel Duffy, Tsengdar Lee and Chaowei Yang. A High Performance Hadoop-based Framework for Handling Big Array-based Spatiotemporal Data

Lot’s more work to be done…

- Continue work with GMU and the spatiotemporal indexing approach
- Comprehensive testing out FUSE with HDFS
- Potential exploration of Hadoop on top of high-performance file systems, such as GPFS or Lustre or Gluster
- Testing of in-memory technologies, such as Spark
- Working with other communities, namely the Earth System Grid Federation (ESGF) to introduce CAaaS into ESGF
  - Intergovernmental Panel on Climate Change (IPCC) Assessment Report (AR) data
  - Observations for Model Intercomparison (OBS4MIPS)
  - Analysis for Model Intercomparison (Ana4MIPS)
Thank you & enjoy the reception!