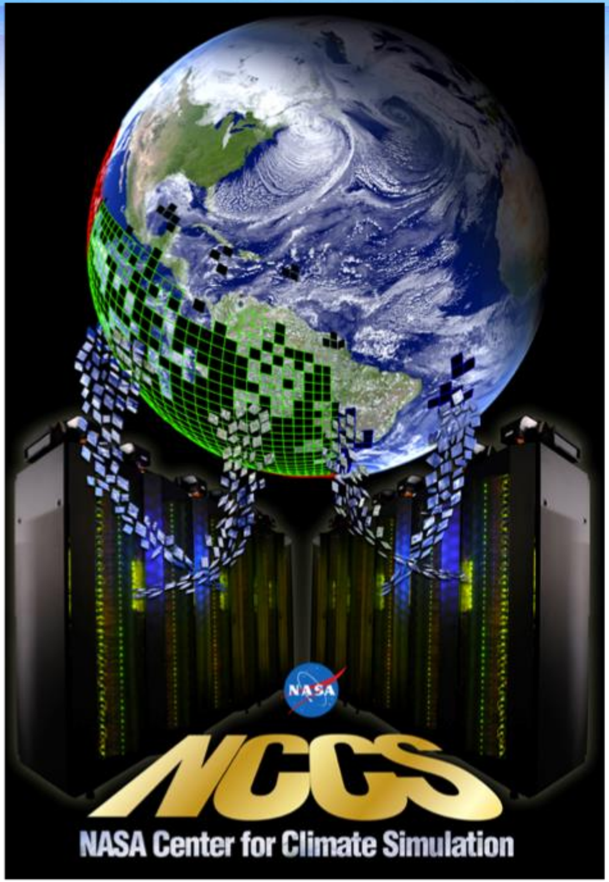


Meeting the Big Data Challenges of Climate Science through Cloud-Enabled Climate Analytics-as-a-Service

Daniel Duffy (daniel.q.duffy@nasa.gov)
NASA Center for Climate Simulation
NASA Godard Space Flight Center

John Schnase (john.schnase@nasa.gov)
Office of Computational and Information Sciences and Technology
NASA Godard Space Flight Center

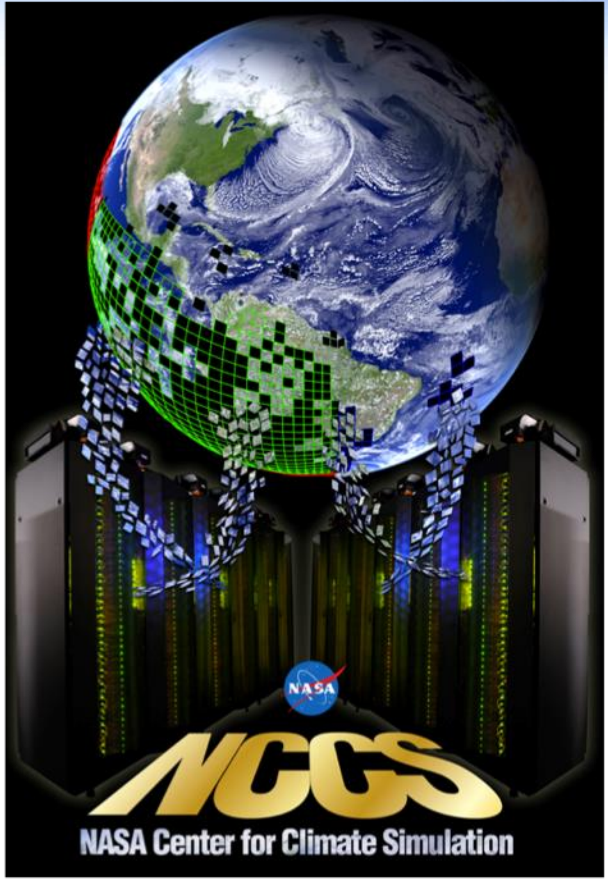
NASA Center for Climate Simulation (NCCS)

Funded by the Science Mission Directorate

- Located at the Goddard Space Flight Center (GSFC)

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

- State-of-the-art high-performance computing, data storage, and networking technologies
- Advanced analysis and visualization environments
- High-speed access to petabytes of Earth Science data
- Collaborative data sharing and publication services



http://www.nccs.nasa.gov

Data Centric HPC, Big Data and IT Environment

2006-

Data Sharing and Publication

- Capability to share data & results
- Supports community-based development
- Data distribution and publishing

Code Development

- Code repository for collaboration
- Environment for code development and test
- Code porting and optimization support
- Web based tools

User Services

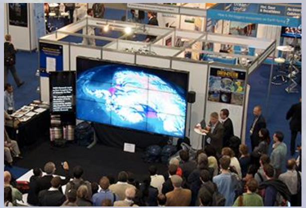
- Help Desk
- Account/Allocation support
- Computational science support
- User teleconferences
- Training & tutorials

DATA Storage & Management

Global file system enables data access for full range of modeling and analysis activities

Analysis & Visualization

- Interactive analysis environment
- Software tools for image display
- Easy access to data archive
- Specialized visualization support



Data Transfer

- Internal high speed interconnects for HPC components
- High-bandwidth to data center users
- Multi-gigabit network supports on-demand data transfers

Security

HPC Computing

- Large scale HPC computing
- Comprehensive toolsets for job scheduling and system monitoring

Data Archival and Stewardship

- Large capacity storage
- Tools to manage and protect data
- Data migration support

NCCS Computational Growth

Continue to deploy scalable units into the Discover Cluster

Truly a hybrid system

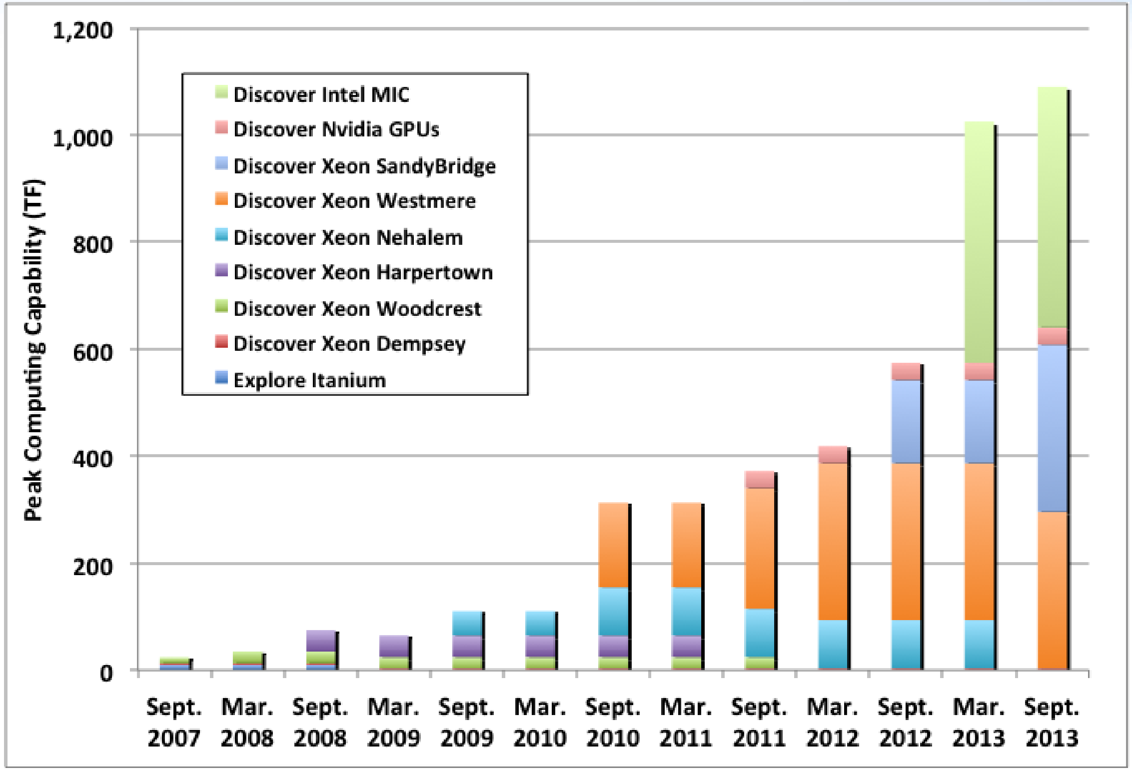
- Xeon only nodes
- Nodes with GPUs
- Nodes with Intel Phi

Major milestone for the NCCS in 2012

- Exceeded 1 PF Peak!

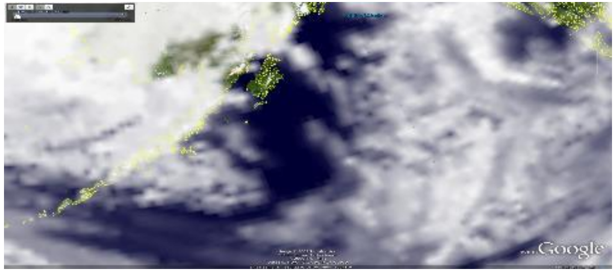
Growth over the last 10 years

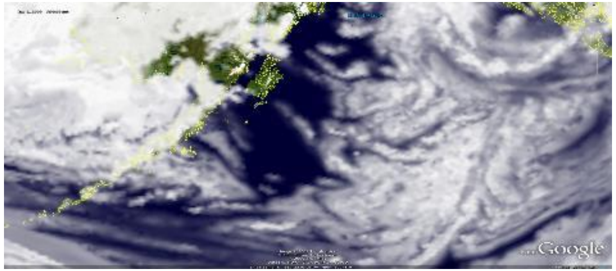
- 300x increase in compute
- 2,000x increase in storage

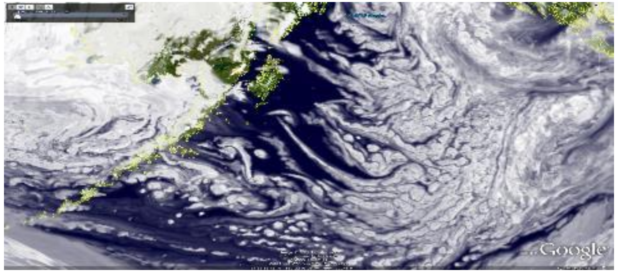


Increasing Global Model Resolution

Current Operations	Cloud-Permitting	Cloud-Resolving







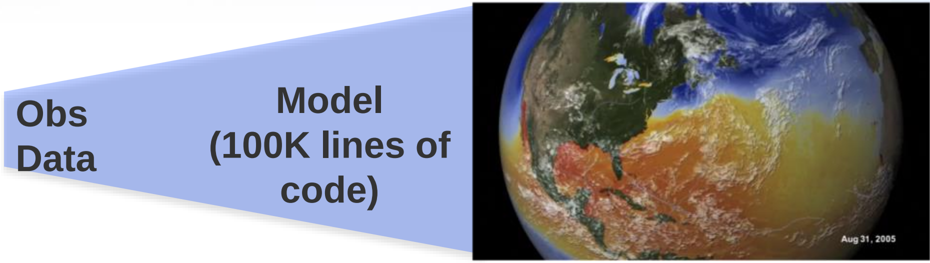
Requirement	Current Operations	Cloud-Permitting	Cloud-Resolving
Number of Cores	100’s	300,000	10,000,000
Resolution	27 KM	10 KM to 3 KM	1 KM or Finer
Number of Racks	1 Rack	234 Racks	7,800 Racks
Total Power	20 KW	4. 7 MW	100 MW

Assuming current compute technology (Intel SandyBridge), the computer needed to run a cloud-resolving model does not exist today and would require entirely too much power. A different approach is needed – adoption of low-power highly parallel processors.

Typical HPC Applications

Takes in small input and creates large output

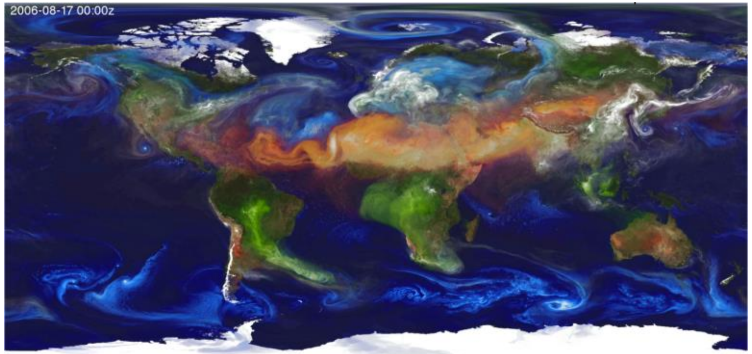
- Using relatively small amount of observation data, models are run to generate forecasts
- Fortran, Message Passing Interface (MPI), large shared parallel file systems
- Rigid environment – users adhere to the HPC systems



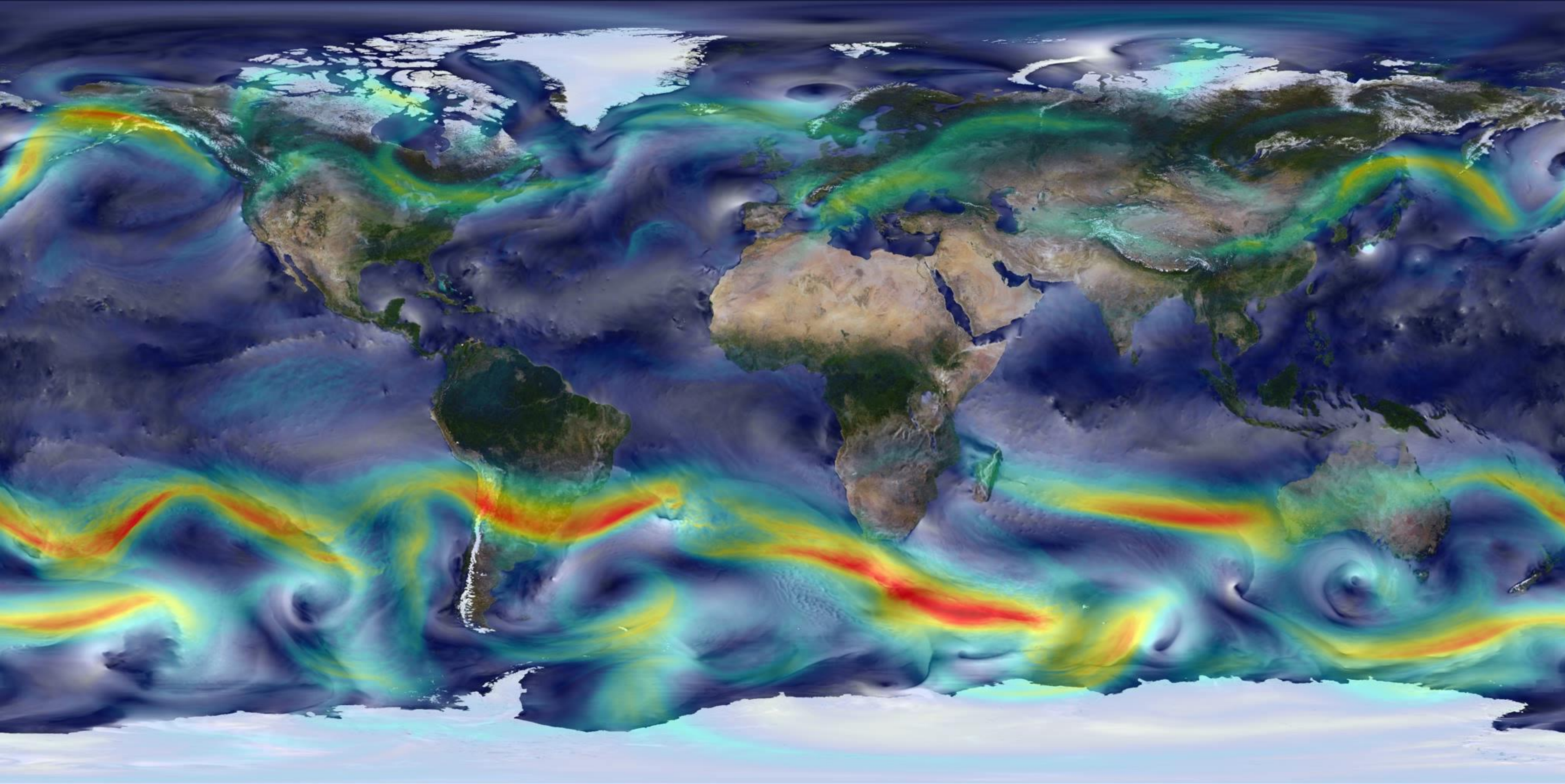
10-km GEOS-5 meso-scale simulation for Observing System Simulation Experiments(OSSEs)

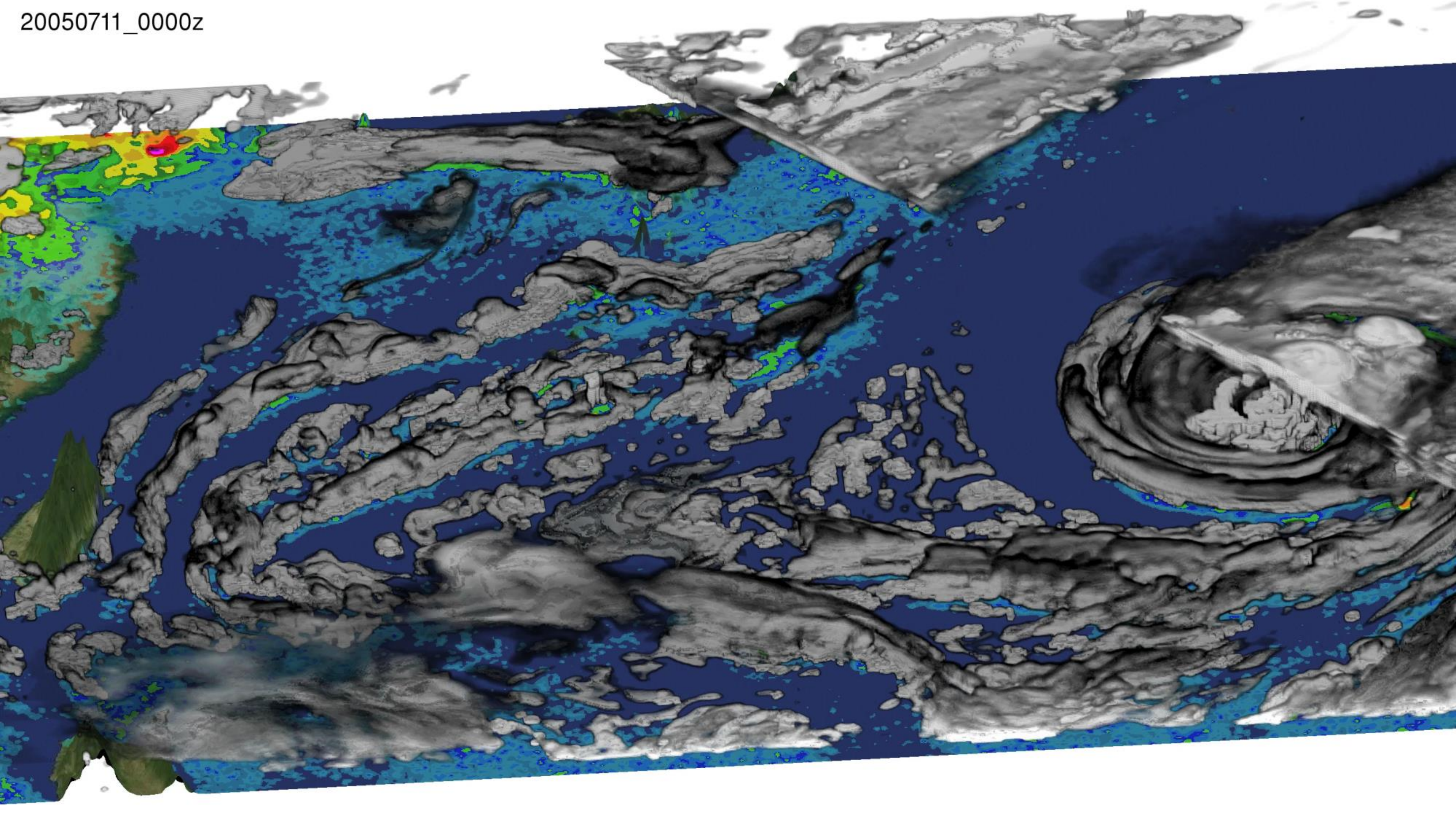
Example: GEOS-5 Nature Run (GMAO)

- 2-year Nature Run at 7.5 KM resolution
- 3-month Nature Run at 3.5 KM resolution
- Will generate 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/



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.




20050711_0000z

Typical Analysis Applications

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

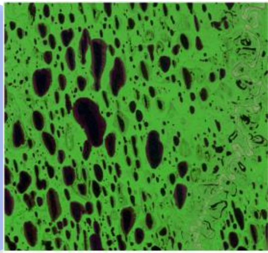
- Using large amounts of distributed observation and model data to generate science
- Python, IDL, Matlab
- Agile environment – users run in their own environments

Examples

- Evaporative transport (Wei experiment)
  - 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

Yukon Delta Alaska; courtesy of Landsat
http://landsat.visibleearth.nasa.gov/view.php?id=72762

Analysis
(100's of lines of code)



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

Planning Science/Proposal Writing

What question am I trying to answer?

- Example: Suppose we want to generate maps of surface water from 1990 to 2012 in the arctic boreal region (problem courtesy of Mark Carroll, Code 618)

What data are available and where are they?

- Landsat time series available at the LP DAAC

How much data is needed?

- Full time series requires >100,000 scenes and ~20TB of data storage

Can I store all that data? If not, how can I process it?

- No. So download chunks of 5TB to local machine. Process. Delete. Download more.
- Projected time – 9 months – without any mistakes!

That's too long, so how can I modify my science question accordingly?

- Average across three epochs (1990, 2000, 2010)
- 25,000 scenes and ~7TB of data
- Projected time – 2 to 3 months

Scientists are limiting their questions (and science) based on the IT resources of their desktops!

Conversations Between Scientists or Conversations "We Don't Want" Between Scientists

Scientist 1

Scientist 2

Scientist 1: Hey, what are you working on these days?

Scientist 2: Oh, you know, just processing data from the new satellite for my ROSES project.

Scientist 1: You know, I need that same data for my project. Where did you get that?

Scientist 2: I downloaded it from the web.

Scientist 1: How long did it take you?

Scientist 2: Quite awhile; several weeks.

Scientist 1: Oh, man, I don't want to have to download all That data and take several weeks. Do you think I could get a copy from you?

Scientist 2: Sure, I am just not sure how to get it to you. I could NFS serve it from my machine to yours or just give you access to my system.

Scientist 1: That would be great. You don't think the security guys would mind do you?

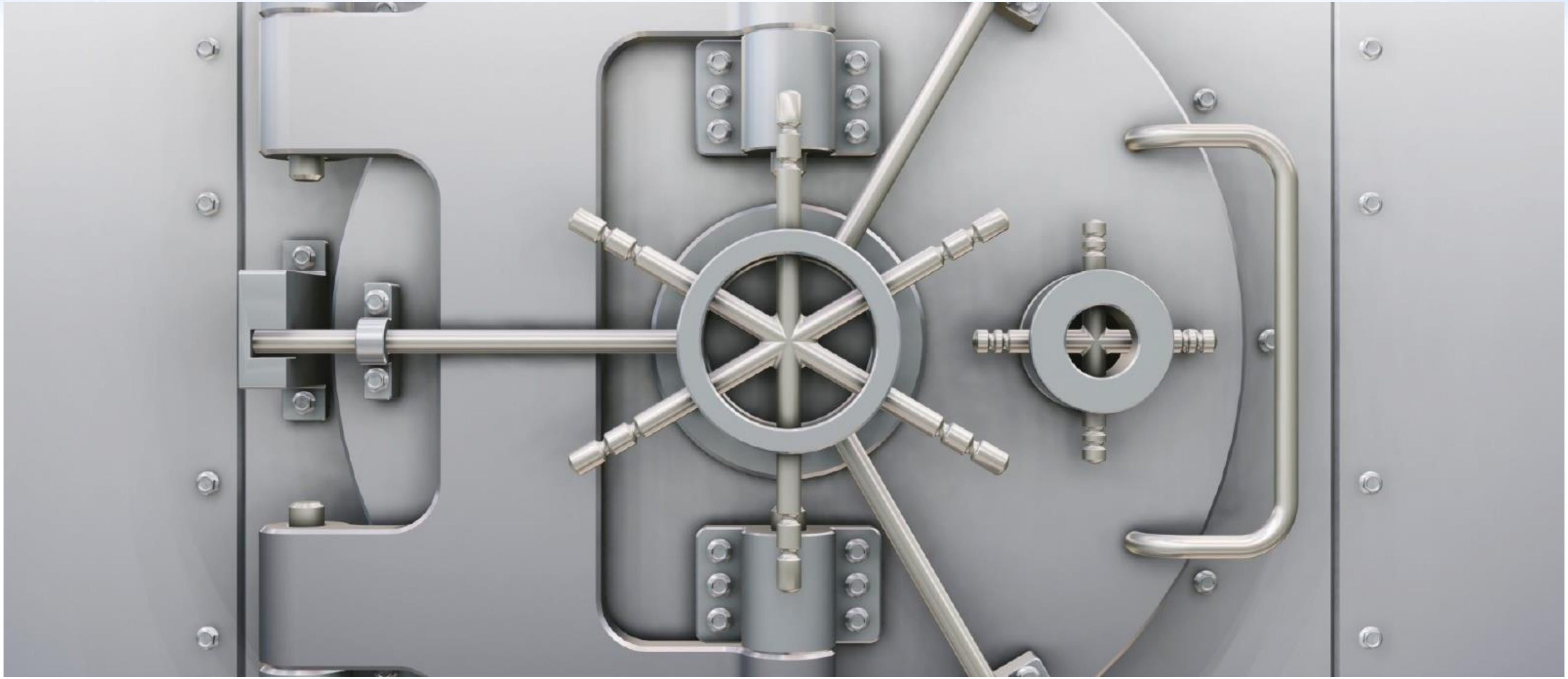
Scientist 2: No, I'm sure they wouldn't. It is in the name of science after all.

Perception of our archive

How do users doing analytics view NASA data systems and archives?

System Administrator’s View of Archive Storage

User's View of Archive Storage



How Do User's Read the Data?

This is how we deliver data

Evolution to a Data Services Centric Environment

Data

HPC Models
- GEOS 5
- ModelE
- WRF

Observations
- Ground Based
- Satellite
- In Situ

Reanalysis
- MERRA
- NOAA
- Others

HPC Computing and Storage
- NASA NCCS
- NOAA
- Others

Analytics

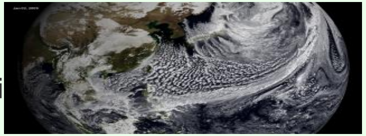
Data Services

Moving beyond just a file system and a storage repository.

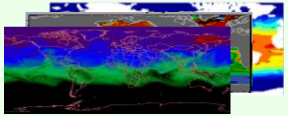
NCCS and Data Services Projects
- Dali Analysis Nodes
- vCDS
- Hadoop (HDFS)
- Merra Analytic Service
- Earth System Grid
- Web Portals

Discovery

Modelers/Scientists



Downstream Users
- Agriculture
- Water Management
- Health
- Famine Prediction



Commercial
- Insurance/Reinsurance
- Commodity Trading

Public/Citizen Scientists

Data Management System
iRODS based management of federated data sets

Data Analysis and Analytics Technology Gap

Archive

Archive
~1 PB of Disk
~35 PB of Tape
Optimized for long term storage, typically slower storage designed for streaming reads and writes

Leads to Un-optimized Practices:
Users perform data analysis straight from the archive and complain that it is too slow.

Very Large Performance Gap

Specifically for Data Analysis, Analytics, and Visualization of large scale data

What technologies can we use to help bridge this gap?

Large Scale Compute

Discover Cluster
>1 PF Peak
~18 PB of Disk
Optimized for large scale simulations with fast storage designed for streaming applications

Leads to Un-optimized Practices:
Users analyze large data sets through a series of many small blocks reads and writes and complain that it is too slow.

Shifting Technologies Toward Big Data

High Performance Computing

- Shared everything environment
- Very fast networks; tightly coupled systems
- Cannot lose data
- Big data (100 PBs)
- Bring the data to the application
- Large scale applications (up to 100K cores)
- Applications cannot survive HW/SW failures
- Commodity and non-commodity components; high availability is costly; premium cost for storage

Object Storage
MapReduce
Hadoop
Cloud
Open Stack
Virtualization
Accelerators

Large Scale Internet

- Examples: Google, Yahoo, Amazon, Facebook, Twitter
- Shared nothing environment
- Slow networks
- Data is itinerant and constantly changing
- Huge data (Exabytes)
- Bring the application to the data
- Very large scale applications (beyond 100Ks)
- Applications assume HW/SW failures
- Commodity components; low cost storage

Very Big Data!

Google

- By 2012, Gmail had 425 million active users[1]
- Each user gets 15 GB of storage for free
- 425,000,000 * 15 GB = 6,375,000,000 GB = 6,385,000 TB = 6,375 PB = 6.375 EB
- Assuming about 6% of the email is spam[2], Gmail carried around 382.5 PB of spam!

Facebook

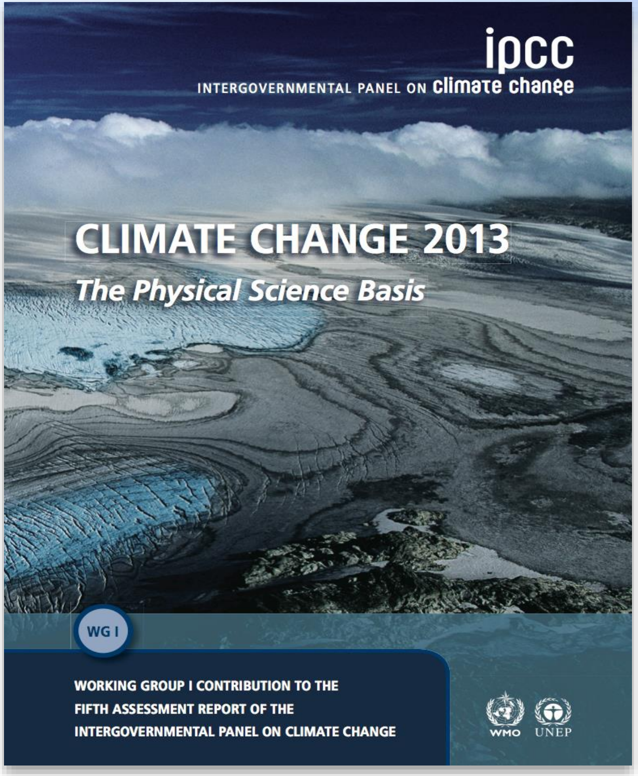
- By 2012, Facebook was processing 500 TB of data per day[3]
- 2.7 billion Like actions and 300 million photos per day
- Facebook scanned about 105 TB every 30 minutes[4]

1. http://venturebeat.com/2012/06/28/gmail-hotmail-yahoo-email-users/
2. http://krebsonsecurity.com/2013/01/spam-volumes-past-present-global-local/
3. http://news.cnet.com/8301-1023_3-57498531-93/facebook-processes-more-than-500-tb-of-data-daily/
4. http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/

How Much Climate Data?

How big?

- MERRA Reanalysis Collection ~200 TB
- Total data holdings of the NASA Center for Climate Simulation (NCCS) is ~40 PB
- Intergovernmental Panel on Climate Change Fifth Assessment Report ~5 PB (data on line now)
- Intergovernmental Panel on Climate Change Sixth Assessment Report ~100 PB (to be created within the next 5 to 6 years)



Our View of “Big Data”

Think friction …

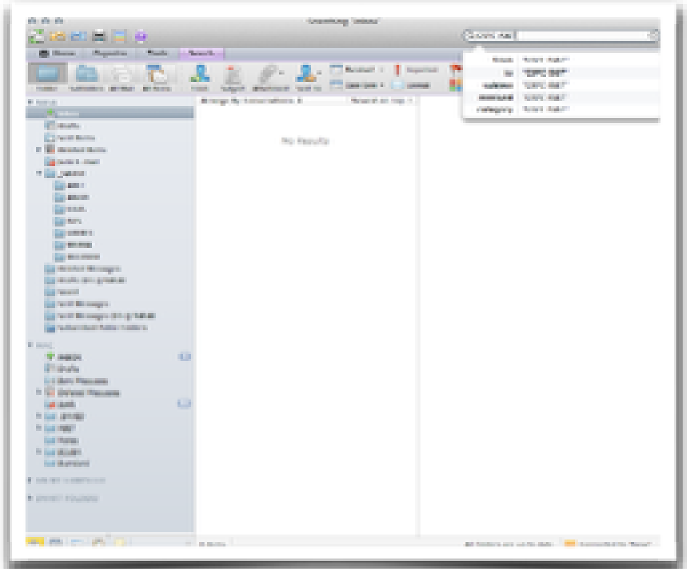
Data bigness depends on ease of use for the type of questions being asked …

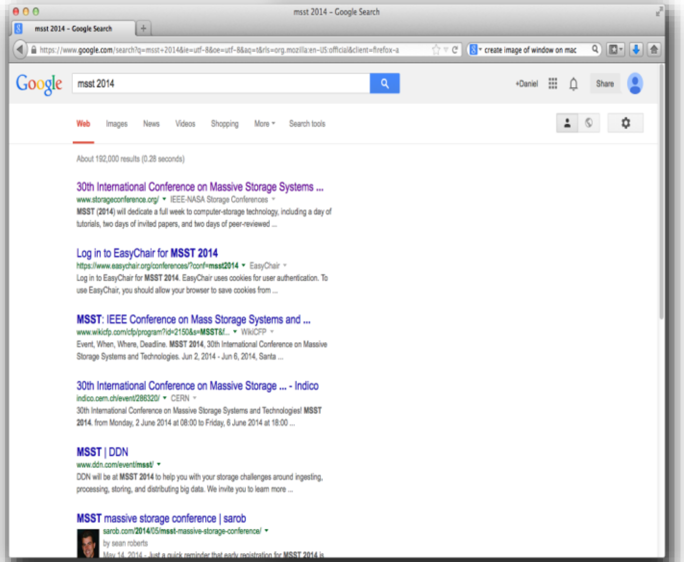
… and a particular technology may or may note help.

Query: “MSST 2014”

Google: 192,000 results in less than 1 sec.
Outlook: No results, about as fast.

Successful interactions with data result when a resonance relationship sets up between data, technology, and ease of use.





Do you have a big data problem?

Now Google the following:

- Analytics – 152,000,000 results in 0.21 seconds
- Cloud Computing – 287,000,000 results in 0.32 seconds
- Garage – 192,000,000 results in 0.37 seconds

Have you ever asked someone to resend you an email that you can no longer find?

- Your email is a “Big Data” problem!

Reference

- “A Vast Machine” by Paul Edwards

What are the Critical Elements for Climate Analytics?

High-Performance Compute/Storage Fabric

Storage-proximal analytics
Canonical operations

Data can't move, analyses need horsepower, and leverage requires something akin to an analytical assembly language …

Data

Relevance
Collocation

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

Exposure

Convenience
Extensible

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

Climate Analytics as a Service

MERRA Reanalysis

Data

Relevance
Collocation

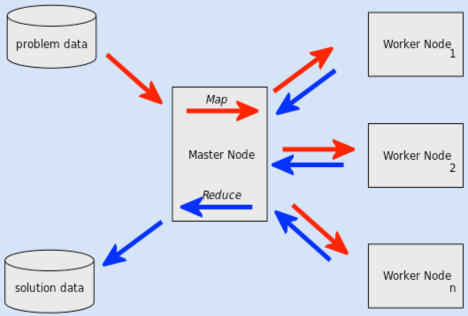
Data have to be significant, sufficiently complex, and physically or logically co-located to be interesting and useful …

High-Performance Compute/Storage Fabric

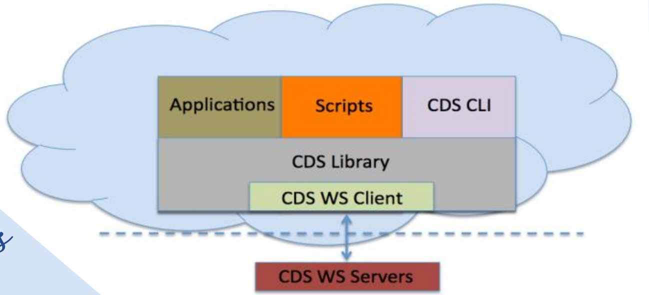
Storage-proximal analytics
Canonical operations

Data can't move, analyses need horsepower, and leverage requires something akin to an analytical assembly language …

MERRA Analytic Services



Climate Data Services API



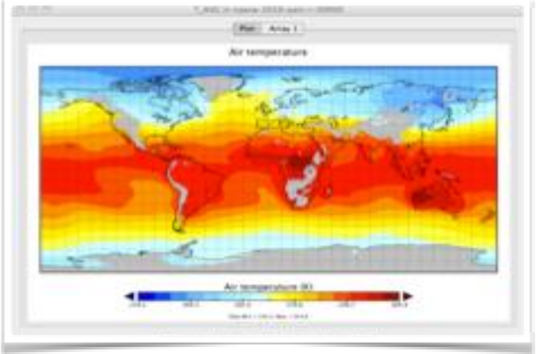
Exposure

Convenience
Extensible

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

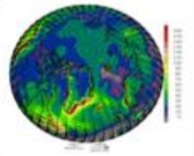
MERRA Data Set

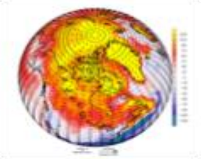
MERRA Reanalysis



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 …





ESGF MERRA published variables

CMIP5	MERRA	Units	Description(Long Name)
rlus	rlus	W m-2	Surface Upwelling Longwave Radiation
rlut	lwtup	W m-2	TOA Outgoing Longwave Radiation
rlutcs	lwtupclr	W m-2	TOA Outgoing Clear-Sky Longwave Radiation
rsds	swgnt	W m-2	Surface Downwelling Shortwave Radiation
rsdscs	swgdnclr	W m-2	Downwelling Clear-Sky Shortwave Radiation
rsdt	swtdn	W m-2	TOA Incident Shortwave Radiation
rsut	swtdn??	W m-2	TOA Outgoing Shortwave Radiation
clt	cldtot	%	Total Cloud Fraction
pr	prectot	kg m-2 s-1	Precipitation
cl	cloud	%	Cloud Area Fraction
evspsbl	evap	kg m-2 s-1	Evaporation
hfls	eflux	W m-2	Surface Upward Latent Heat Flux
hfss	hflux	W m-2	Surface Upward Sensible Heat Flux
hur	rh	%	Relative Humidity
hus	qv	v	Specific Humidity
prc	preccon	kg m-2 s-1	Convective Precipitation
prsn	precsno	kg m-2 s-1	Snowfall Flux
prw	tqv	kg m-2	Water Vapor Path
ps	ps	Pa	Surface Air Pressure
psl	slp	Pa	Sea Level Pressure
rlds	lwgnt	W m-2	Surface Downwelling Longwave Radiation
rldscs	lwgabclr	W m-2	Surface Downwelling Clear-Sky Longwave Radiation
rsutcs	swtdn	W m-2	TOA Outgoing Clear-Sky Shortwave Radiation
ta	t	K	Air Temperature
tas	t2m	K	Near-Surface Air Temperature
tauu	taux	Pa	Surface Downward Eastward Wind Stress
tauv	tauy	Pa	Surface Downward Northward Wind Stress
tro3	o3	1.00E-09	Mole Fraction of O3
ts	ts	K	Surface Temperature
ua	u	m s-1	Eastward Wind
uas	u10m	m s-1	Eastward Near-Surface Wind
va	v	m s-1	Northward Wind
vas	v10m	m s-1	Northward Near-Surface Wind
wap	omega	Pa s-1	omega (=dp/dt)
zg	h	m	Geopotential Height

MERRA Analytics Service (MERRA A/S)

MapReduce

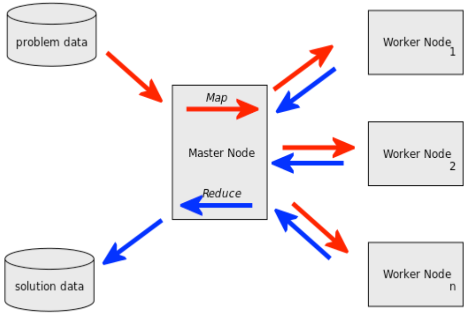
- MapReduce is a framework for processing parallelizable problems across huge datasets using a large number of computers.
- Computational processing can occur on data stored either in a filesystem (unstructured) or in a database (structured).
- MapReduce can take advantage of locality of data, processing data on or near the storage assets to decrease transmission of data.
- "Map" step: The master node takes the input, divides it into smaller sub-problems, and distributes them to worker nodes. A worker node may do this again in turn, leading to a multi-level tree structure. The worker node processes the smaller problem, and passes the answer back to its master node.
- "Reduce" step: The master node then collects the answers to all the sub-problems and combines them to form the output – the answer to the problem it was originally trying to solve.

Much of the MapReduce work has been building the code ecosystem to manage multidimensional binary NetCDF files …

Cluster

- 36 node Dell cluster, 576 Intel 2.6 GHz SandyBridge cores, 1300 TB raw storage, 1250 GB RAM, 11.7 TF theoretical peak compute capacity.
- FDR Infiniband network with peak TCP/IP speeds >20 Gbps.

MERRA Analytic Services



Canonical Ops Library

- We're also creating 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, avg, max, min, var, sum, count operations of the general form:

$$result \;<=\; \mathrm{avg}(var,\ (t_0,t_1),\ ((x_0,y_0,z_0),(x_1,y_1,z_1))),$$

that return, in this example, the average of a variable when given a variable name, temporal extent, and spatial extent ...

- Averages over time, space, and elevation can be performed now for all MERRA variables.

Hadoop File System Organization

- Total size of the native, compressed NetCDF MERRA collection in a standard filesystem ~80 TB.
- Native MERRA files are sequenced and ingested into the Hadoop cluster in triplicated 640 MB blocks.
- Total size of MERRA/AS HDFS repository ~480 TB.

5621 lines of MapReduce code behind avg operation …

Climate Data Services Application Programming Interface (CDS-API)

CDS Reference Model

Ingest – Submit/register a Submission Information Package (SIP).

Query – Retrieve data from a pre-determined service request (synchronous).

Order – Request data from a pre-determined service request (asynchronous).

Download – Retrieve a Dissemination Information Package (DIP).

Status – Track progress of service activity.

Execute – Initiate a service-definable extension. Allows for parameterized growth without API change.

CDS Library

```
Class CDSLibrary(object):

    def order(self, service, parms):
            cds_ws.order(service, parms)

    def avg(self, service, parms, destination):
            sessionId = cds_ws.order(service, parms)
            response = cds_ws.status(service, sessionId)
        ....... Loop until result is available
            cds_ws.download(service, sessionId, destination)
```

CDS CLI

```
Welcome to the NASA GSFC CISTO Climate Data Services (cds).
Type help or ? to list commands.

(nasa-gsfc-cisto-cds) order MAS parms!
GetAverageByVariable_TimeRange_SpatialExtent_VerticalExtent
&operation=avg&variable_list=T&start_date=201101&end_date=201102&a
vg_period=2&min_lon=-125&min_lat=24&max_lon=-66&max_lat=50&start
_level=13&end_level=13'

(nasa-gsfc-cisto-cds) execute HADOOP mapreduce jar!/opt/cds/bin/cds-
mas-mapreduce.jar  inputPath!/opt/cds/seq-input/merra/2011 outputPath!/
opt/cds/merra_2011_mr_seqout/npana
```

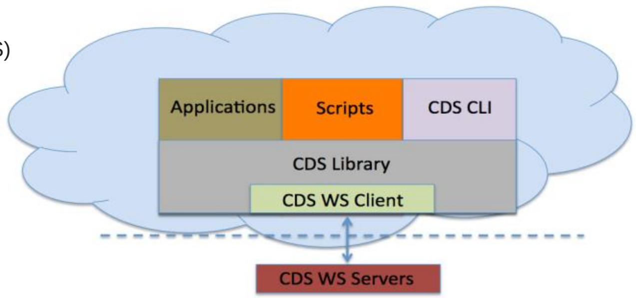
CDS Client Stack

- The MERRA/AS project has been the starting point for development of the NASA Climate Data Services (CDS) Application Programming Interface (API).
- The CDS client stack can be distributed as a software package or used to build a cloud service (SaaS) or distributable cloud image.
- This approach to API design focuses on the specific analytic requirements of the climate sciences and marries the language and abstractions of collections management (OAIS) with those of high-performance analytics (MapReduce) ...

CDS Applications

```
[gtamkin@localhost python]$ more ./user_app_ext.py
from cds import CDSApi
cds_api = CDSApi()

service = 'MAS'
north_american_parms =
'GetAverageByVariable_TimeRange_SpatialExtent_VerticalExtent
&operation=avg&variable_list=T&start_date=201101&end_date=201102&a
vg_period=2&min_lon=-125&min_lat=24&max_lon=-66&max_lat=50&start
_level=13&end_level=13'
destination=/home/gtamkin/avg-out'

Class UserAppExt(object):

if __name__ == '__main__':
    sessionId = cds_api.avg(service, north_american_parms, destination)
    print "processing complete for = " + filename
```

Climate Data Services API



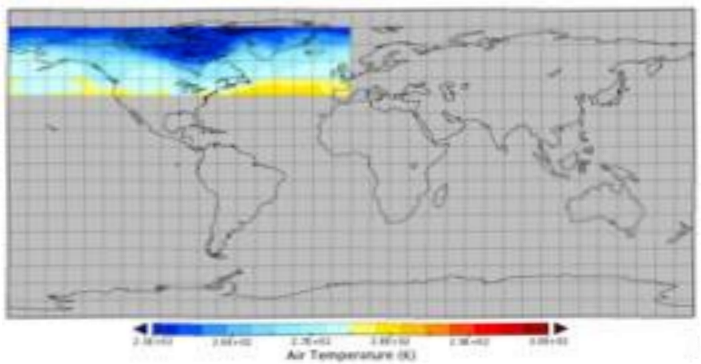
CDS Scripts

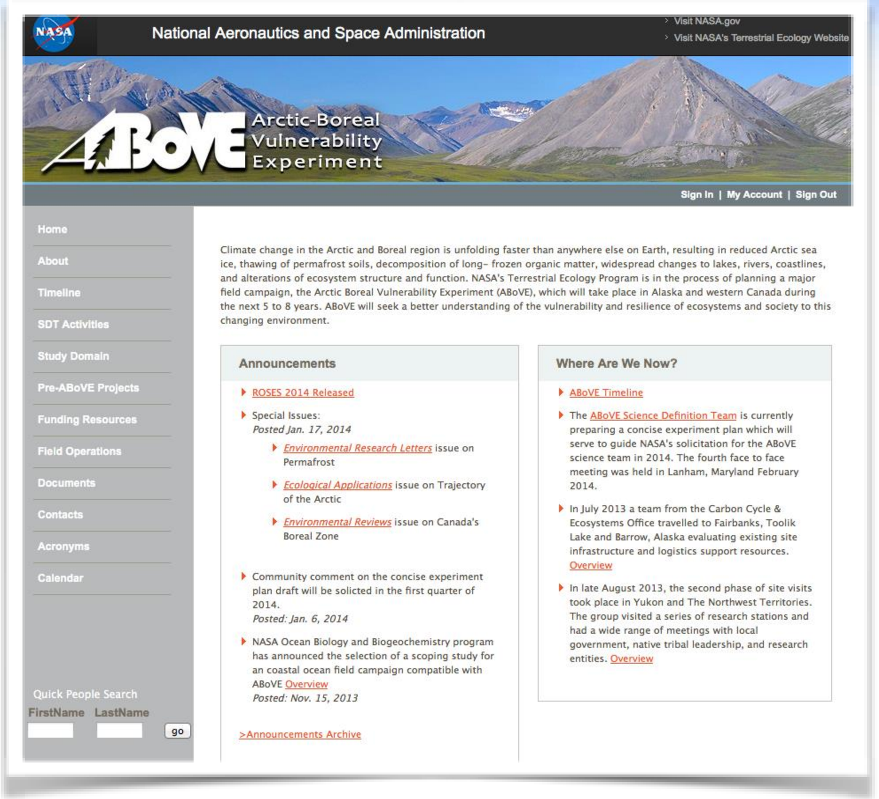
```
#!/usr/bin/env python
import time

from CDSLibrary import CDSApi
from wei_input import WEIInput
wei_exp = WEIInput()

# The rest of the file is run by the Python interpreter.
__doc__ = """This string is treated as the module docstring."""

service = wei_exp.getService()
catalog = wei_exp.getInput()
destination = wei_exp.getDestination()

cds_lib = CDSApi()
logger = cds_lib.getLogger()

start_time = time.time()

logger.debug("Generating: ca_avg_temp")
input = cds_lib.encode(catalog["ca_avg_temp_dictionary"])
cds_lib.avg(service, input, destination)

exit()
```

Where is the resonance with science?

- Air Temperature, Precipitation / Avg, Max, Min / 1979-2014 / monthly means, 3-hourly
- Traditional: Find and order from archive (hrs?)
  Transfer ~100 GB (~1 hr, depending)
  Client-side clip/compute using GrADS
  1-1.5 days
- MERRA/AS: Server-side clipping using OPeNDAP (single stream op, time ??, > 2 mos)
  Server-side clip/compute (~24 hrs)
  Transfer final product ~1.5 GB

Takes about as long, but the scientist is free to work on other things …

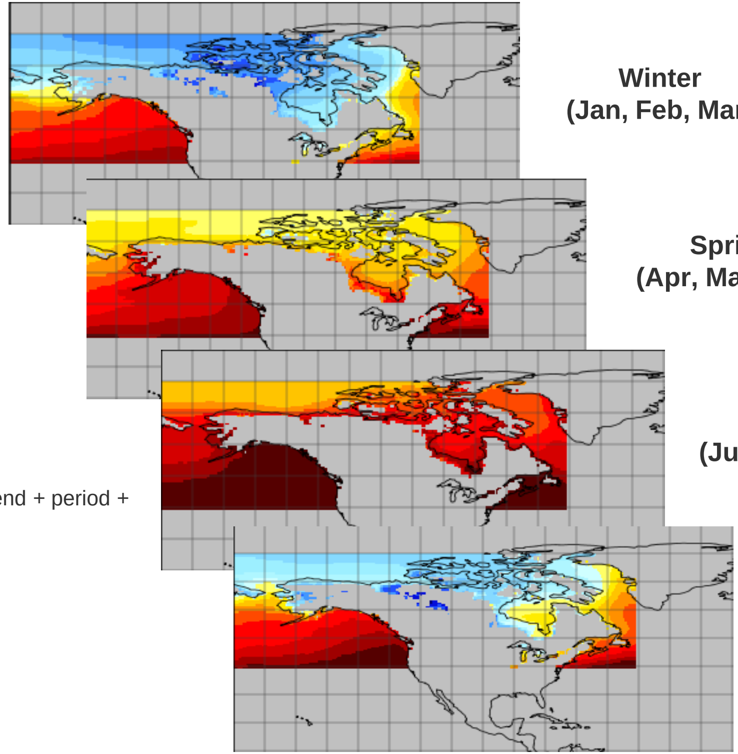




Simple ABoVE Related Example

QUESTION: Extract the average temperature by season for the year 1980 for the ABoVE region at every vertical height in the MERRA data.

```
#!/usr/bin/env python
"""
Created on February 6, 2014

@author: dqduffy
"""
import sys
from CDSLibrary import CDSApi
cds_lib = CDSApi()
service = "MAS"

name       = "above_avg_seasonal_temp_1980_instM_3d_ana_Np"
job        = "&job_name=" + name
collection = "&collection=instM_3d_ana_Np"
request    = "&request=GetVariableBy_TimeRange_SpatialExtent_VerticalExtent"
variable   = "&variable_list=T"
operation  = "&operation=avg"
start      = "&start_date=198001"
end        = "&end_date=198012"
period     = "&avg_period=3"
space      = "&min_lon=-180&min_lat=40&max_lon=-50&max_lat=80"
levels     = "&start_level=1&end_level=42"
file_job_epoch1_aveT  = "./" + name + ".nc"
above_job_epoch1_aveT = job + collection + request + variable + operation + start + end + period +
space + levels

class UserApp(object):

  if __name__ == '__main__':

      # exercise all canonical operations
      print(above_job_epoch1_aveT)
      cds_lib.avg(service, above_job_epoch1_aveT, file_job_epoch1_aveT)
```



Winter
(Jan, Feb, Mar)

Spring
(Apr, May, Jun)

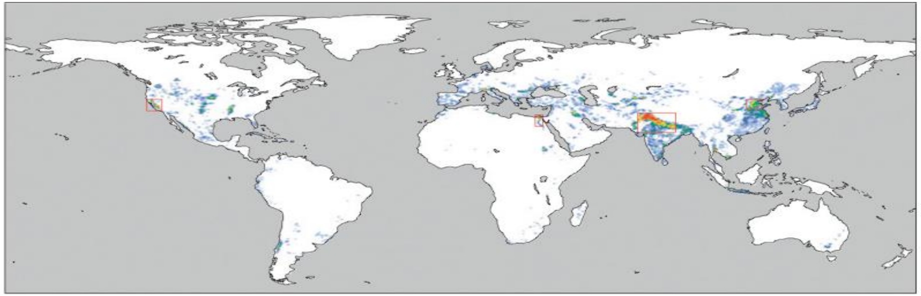
Summer
(Jul, Aug, Sep)

Fall
(Oct, Nov, Dec)

Wei Experiment

An Estimation of the Contribution of Irrigation to Precipitation Using MERRA

- Wei team used MERRA data to study four intensively irrigated regions: *northern India/Pakistan*, the North China Plain, the California Central Valley, and the Nile Valley.
- Seasonal rates of evapotranspiration with and without irrigation over the studied areas were then compared to assess the impact of irrigation.
- The data required for these calculations include precipitation, evapotranspiration, temperature, humidity, and wind at different tropospheric levels at six-hourly time steps from 1979 to 2002.
- This early-stage data reduction—average values for environmental variables over specific spatiotemporal extents—is the type of data assembly that historically has been performed on the scientist's workstation after transfers from public archives of large blocks of data.



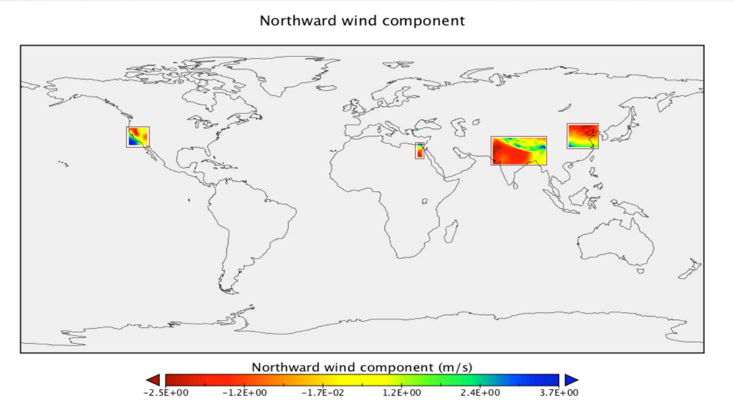
THE UNIVERSITY OF TEXAS AT AUSTIN

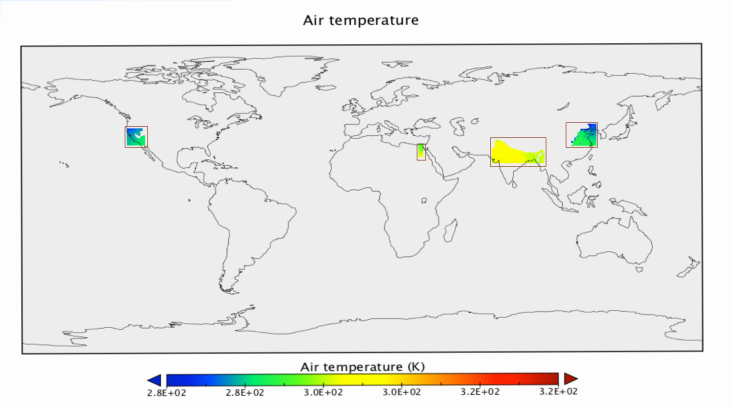
JACKSON SCHOOL OF GEOSCIENCES

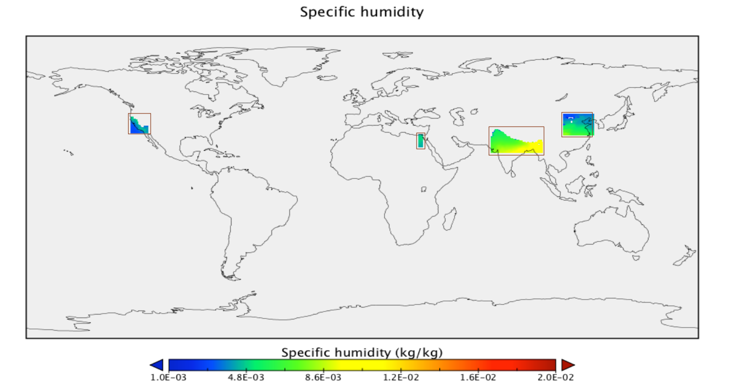
Wei, J., Dirmeyer, P. A., Wisser, D., Bosilovich, M. G., & Mocko, D. M. (2013). Where does irrigation water go? An estimate of the contribution of irrigation to precipitation using MERRA. *Journal of Hydrometeorology*, 14(2), 271–289.

Wei Experiment

An Estimation of the Contribution of Irrigation to Precipitation Using MERRA







Wei, et al.

- ~8.4 TB transferred from archive to local workstation (weeks)
- Clipping, averaging performed by Fortran program on local workstation (days)

MERRA/AS (Time trials in progress …)

- Clipping, averaging performed by MERRA/AS (less than one day)
- ~35 GB of final product moved to local workstation

- Significant time savings in data wrangling,

- Rapid screening over monthly means files takes minutes, and

- Possibility of folding Dr. Wei's modeling algorithm back into the CDS API …

What is Climate Analytics?

Analytics is the discovery of meaningful patterns in collections of data.

Climate **Climate**

Climate Analytics is the discovery of meaningful patterns in collections of climate data.

large

Climate Analytics is the discovery of meaningful patterns in large collections of climate data.

Climate Analytics-as-a-Service

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).

MERRA Analytic Services (MERRA/AS) is an example of CAaaS. MERRA/AS enables MapReduce analytics over NASA's Modern-Era Reanalysis for Research and Applications (MERRA) reanalysis dataset. MERRA/AS's capabilities are delivered to applications and customers through NASA's Climate Data Services API.

Next Steps

High Performance

- Continued development of the CDS-API
- Roll out for beta testers and version 1 deployment
- Extension of the CDS-API to persistent services and create a generative environment for the API
- High Performance Science Cloud (virtualization)

Continue to Expand Science Support

- NCCS Data Portal
- ABoVE Mission Support
  Hadoop Cluster as a Service
- Nature Run Data Processing
- Looking for Other Science Opportunities

Special Thanks

Management and Leadership

- Phil Webster
- John Schnase
- Mark McInerney

The People Who Actually Make Things Work

- Scott Sinno
- Hoot Thompson
- Garrison Vaughan
- Al Settell

The Scientists

- Mark Carroll
- Peter Griffith
- Tatiana Loboda
- Dr. Elizabeth Hoy

THANK YOU

And Thank You to the All People That Make This Work!