Atmospheric Science Data Center Update

CERES Science Team Meeting
26 April 2011
John M Kusterer
Head, ASDC
• Update on CERES Ordering Metrics
• Status of CERES Production Environment
• CERES Data Access
  • Subsetting
  • Web pages
  • Future vision of data access
Update on CERES Ordering Metrics
CERES and FLASHFlux Archive Volume

By Data Year

Total Volume Archived ~560TB

Data Volume (TB)


TERRA
AQUA
FLASH

April 2011 ASDC Update for CERES STM Page: 4
CERES and FLASHFlux Data Distribution

Total data distributed (thru March 31, 2011)
- ~485 TB
- 15451 orders
- 2,779 different users

Data Volume (TB)

<table>
<thead>
<tr>
<th>Year</th>
<th>Data Volume (TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0</td>
</tr>
<tr>
<td>2001</td>
<td>0</td>
</tr>
<tr>
<td>2002</td>
<td>0</td>
</tr>
<tr>
<td>2003</td>
<td>0</td>
</tr>
<tr>
<td>2004</td>
<td>0</td>
</tr>
<tr>
<td>2005</td>
<td>0</td>
</tr>
<tr>
<td>2006</td>
<td>0</td>
</tr>
<tr>
<td>2007</td>
<td>0</td>
</tr>
<tr>
<td>2008</td>
<td>0</td>
</tr>
<tr>
<td>2009</td>
<td>0</td>
</tr>
<tr>
<td>2010</td>
<td>0</td>
</tr>
<tr>
<td>2011</td>
<td>0</td>
</tr>
</tbody>
</table>

Legend:
- **TERRA**
- **AQUA**
- **FLASH**
ASDC CERES Data Users

Number of Customers

- 1 to 10
- 25 to 40
- ~60
- ~130
- ~890

April 2011 ASDC Update for CERES STM
Status of CERES Production Environment
• ASDC Modernization through Integration
• New (relatively) processing and storage based on IBM technologies
• Challenges with initial setup and configuration
  • No one person has all the right answers
  • Consultants, vendors experts, in-house expertise
    • Convergence was difficult
• Stability issues primary concern
• Orchestrated a redesign effort within constraints
• Technical review early April
  • Extremely valuable in validating design
• AMI has been very stable recently
Status of CERES Processing Environment

33,880 batch jobs run by SCF users on AMI using Sun Grid Engine in last 3 weeks

<table>
<thead>
<tr>
<th>Date</th>
<th>x86 Processor</th>
<th>Power 6 Processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/04</td>
<td>2,359</td>
<td>2,155</td>
</tr>
<tr>
<td>04/05</td>
<td>2,117</td>
<td>2,531</td>
</tr>
<tr>
<td>04/06</td>
<td>2,415</td>
<td>1,823</td>
</tr>
<tr>
<td>04/07</td>
<td>1,951</td>
<td>1,918</td>
</tr>
<tr>
<td>04/08</td>
<td>1,053</td>
<td>92</td>
</tr>
<tr>
<td>04/09</td>
<td>270</td>
<td>245</td>
</tr>
<tr>
<td>04/10</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>04/11</td>
<td>427</td>
<td>296</td>
</tr>
<tr>
<td>04/12</td>
<td>1,355</td>
<td>1,391</td>
</tr>
<tr>
<td>04/13</td>
<td>3,255</td>
<td>2,469</td>
</tr>
<tr>
<td>04/14</td>
<td>590</td>
<td>226</td>
</tr>
<tr>
<td>04/15</td>
<td>2,469</td>
<td>590</td>
</tr>
<tr>
<td>04/16</td>
<td>1,355</td>
<td>3,255</td>
</tr>
<tr>
<td>04/17</td>
<td>427</td>
<td>296</td>
</tr>
<tr>
<td>04/18</td>
<td>1,391</td>
<td>1,355</td>
</tr>
<tr>
<td>04/19</td>
<td>1,954</td>
<td>1,275</td>
</tr>
<tr>
<td>04/20</td>
<td>1,294</td>
<td>1,309</td>
</tr>
<tr>
<td>04/21</td>
<td>1,275</td>
<td>1,309</td>
</tr>
<tr>
<td>04/22</td>
<td>1,309</td>
<td>1,275</td>
</tr>
<tr>
<td>04/23</td>
<td>1,294</td>
<td>1,275</td>
</tr>
<tr>
<td>04/24</td>
<td>1,309</td>
<td>1,275</td>
</tr>
</tbody>
</table>
Status of CERES Processing Environment

875,556 batch jobs run by SCF users on AMI using Sun Grid Engine in last 12 months

<table>
<thead>
<tr>
<th>Month</th>
<th>x86 processor</th>
<th>Power 6 processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr</td>
<td>1,390</td>
<td>46,789</td>
</tr>
<tr>
<td>May</td>
<td>12,586</td>
<td>32,752</td>
</tr>
<tr>
<td>Jun</td>
<td>32,899</td>
<td>19,680</td>
</tr>
<tr>
<td>Jul</td>
<td>6,877</td>
<td>46,487</td>
</tr>
<tr>
<td>Aug</td>
<td>3,320</td>
<td>22,186</td>
</tr>
<tr>
<td>Sep</td>
<td>10,853</td>
<td>14,340</td>
</tr>
<tr>
<td>Oct</td>
<td>2,873</td>
<td>45,399</td>
</tr>
<tr>
<td>Nov</td>
<td>107,379</td>
<td>29,328</td>
</tr>
<tr>
<td>Dec</td>
<td>76,037</td>
<td>45,595</td>
</tr>
<tr>
<td>Jan</td>
<td>5,114</td>
<td>64,119</td>
</tr>
<tr>
<td>Feb</td>
<td>28,335</td>
<td>46,385</td>
</tr>
<tr>
<td>Mar</td>
<td>81,566</td>
<td>91,642</td>
</tr>
</tbody>
</table>

April 2011 ASDC Update for CERES STM Page: 13
CERES Data Access
• ASDC hosting current CERES Level 3 subsetter
  • Well received by the science community
• Current collaborative effort underway to develop CERES Level 2 subsetting capability
  • Development model has been rapid prototyping (iterative)
  • Progress to date
    • Subsetter developed for SSF Level 2 Edition 3A data
    • Subsetter framework (facilitates interaction between user interface and data subsetter), data subsetter, and final distribution is complete
    • Subsetter’s spatial metadata database contains more complete and accurate footprints of each CERES granule than anything previously available
    • Subsetted files can be delivered in either HDF4 or NETCDF
    • Multiple NETCDF subset results can be bundled into a single file
CERES Data Comparison Plots

- New spatial metadata mined directly from the CERES hdf granules to use in new CERES subsetter
- Default spatial metadata supplied with the CERES granules in the .met files
New spatial metadata mined directly from the CERES hdf granules to use in new CERES subetter

Default spatial metadata supplied with the CERES granules in the .met files
CERES Plots

- New spatial metadata mined directly from the CERES hdf granules to use in new CERES subsetter
- Default spatial metadata supplied with the CERES granules in the .met files

CERES SSF Terra-FM1-MODIS Edition3A

April 2011

ASDC Update for CERES STM

Page: 18
• Current ASDC CERES landing pages and order tools were developed years ago
  • Largely static HTML with simple pull downs
• Working with CERES science team to improve the user experience
  • Plan to leverage web page design supporting CERES subsetters
    • Developed pages have been customer driven
    • Need to ensure that all customer needs are addressed
• Planning long-term development and support models
  • Agreements on interfaces between user interfaces and subsetting/ordering
  • Availability, maintenance, user support
Future: Understanding Benefits of our Data

• Instrument Science Teams create the data and use it themselves
  • Quality checks
  • Algorithm improvements
  • Inter-calibration

• Other Instrument Science Teams
  • Data Fusion with their own instrument output
  • Inter-calibration of similar measurements
  • Evaluation/Comparison of Alternate Algorithms

• Climate Modeling (Global, regional and thematic studies)

• Climatology and Assimilation Models

• AI tools can help make unexpected correlations of measurements
  • Non-obvious relationship analysis
  • Edge Detection algorithms

• Justification for future instrument missions
  • Based on volume/diversity of use

• Data Stewardship investments
  • Based on volume of access and “value” of the data to the community
Future: Net-centric Ops (NCO) Applied to ASDC

- NCO = Information sharing and re-use
  - Involves discovery, accessibility and understanding
  - Many paths to the same data
    - Keep existing ordering capabilities and add more pathways
    - Instrument Team Defined Storage Model based on instrument output
    - Access model does not have to be dependent upon the storage model
  - Data must be convenient to the researcher-users
- Machine-to-machine interface is critical to practical use of “big data”
  - Minimize manual handling of the data to use it
  - Must convey restrictions, boundary conditions, error bands
  - ASDC MUST make Recipients comfortable that ASDC will guarantee long-term availability for future review
    - Data Stewardship Means Never HAVING to say You’re Sorry
Notional Webservices View

Central Discovery Services
- Registry
- MetaData
- Digital Lib
- End-user Browser
- GPU Visualization System
- Code/Command Client
- IDL Client
- Matlab Client
- GIIEP Client
- Reasoning & AI Tools

Central ID Services
- Authentication

Compute & Aggregation Services
- NOAA 4D Weather Cube server
- Portal or Aggregation server
- GIIEP server
- Climate Modeling System

ASDC Data
- CALIPSO Subsetter/Order (Web-based)
- MCSR Subsetter/Order (Web-based)
- CALIPSO Subsetter/Order (Web-based)
- CALIPSO Browse Images (Web-based)

Local Desktop
- dbms
- Instrument Data Fusion +web server

ASDC Order Tool
- (Web-based)

ASDC Update for CERES STM
Authorization
Authentication
Notional Webservices View
Portal or Aggregation server
GIIEP Client
Reasoning & AI Tools
End-user Browser
GPU Visualization System
Code/Command Client
IDL Client
Matlab Client
GIIEP Client
Webservices data flow
Web Page data path
Common data path
Registry
MetaData
Digital Lib
Web server
4/18/11
NOAA 4D Weather Cube server
Climate Modeling System
GIIEP server
22
Conclusion

- Metrics indicate the ASDC is healthy in supporting its mission
- AMI has become more stable, more effective and customers are utilizing its capability
- Subsetting efforts are moving forward and are responsive to user needs
- Effort to improve the user experience for those visiting ASDC data pages is underway
- Strategic planning is underway to make data available to customers in new paradigms