CERES Data Management Group Status Report

CERES Science Team Meeting

September 1st, 2015

Jonathan Gleason
Overview

• Recent DMT Activities
• Current Data Availability
• Production Status Reporting
• Systems Status
• NPP Data Flow Changes to EDOS
• CATALYST v2.0 Status
Data Management Function

Science Working Groups
- Derive & Refine algorithms
- Validate algorithms
- Validate CERES data sets
- Write Quality Summary

Data Management Team
- Implement algorithms (WG Embedded)
- Maintain CERES Software Framework
- Software Testing & Validation
- Production Planning & Scheduling
- Config., Mgmt., & Doc. support

Atmospheric Sciences Data Center
- Ingest data
- Place operational software in production
- Process, Archive & Distribute data sets
- Provide User Services

Algorithm Development → Algorithm Implementation → Data Processing

Currently ~1M production LOC contained in 54 active PGEs
## CERES Production Software Overview

<table>
<thead>
<tr>
<th>Subsystem Number</th>
<th>Subsystem Name</th>
<th>Number of PGEs</th>
<th>LOC (to nearest 1K)</th>
<th>Publicly Available Data Products</th>
<th>Product Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Instrument/Pre-Processor</td>
<td>2</td>
<td>9K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Instrument</td>
<td>6 + lib</td>
<td>162K</td>
<td>BDS</td>
<td>1/day</td>
</tr>
<tr>
<td>2</td>
<td>ERBE-like/ Inversion</td>
<td>5 + lib</td>
<td>31K</td>
<td>ES-8</td>
<td>1/day</td>
</tr>
<tr>
<td>3</td>
<td>ERBE-like/ TSA</td>
<td>2</td>
<td>14K</td>
<td>ES-9, ES-4</td>
<td>1/month</td>
</tr>
<tr>
<td>4.1 – 4.4</td>
<td>Clouds/VIIRS Subset Code</td>
<td>1</td>
<td>21K</td>
<td></td>
<td>12/hour</td>
</tr>
<tr>
<td>4.1 – 4.4</td>
<td>Clouds</td>
<td>11</td>
<td>358K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.5 – 4.6</td>
<td>Inversion</td>
<td>11</td>
<td>110K</td>
<td>SSF</td>
<td>1/hour</td>
</tr>
<tr>
<td>5</td>
<td>SARB</td>
<td>1</td>
<td>36K</td>
<td></td>
<td>1/hour</td>
</tr>
<tr>
<td>6 &amp; 9</td>
<td>TISA-Gridding</td>
<td>7</td>
<td>34K</td>
<td>SSF1deg-Hour, ISCCP-D2like-Day/Nit</td>
<td>60/month, 36/month, 1/month</td>
</tr>
<tr>
<td>11</td>
<td>GGE0</td>
<td>1</td>
<td>6K</td>
<td>ISCCP-D2like-GEO</td>
<td>1/month</td>
</tr>
<tr>
<td>7.2</td>
<td>Synoptic SARB</td>
<td>1</td>
<td>47K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.1 &amp; 8 10</td>
<td>TISA-Averaging</td>
<td>4</td>
<td>102K</td>
<td>SSF1deg-Day, SSF1deg-Month, SYN1deg-(3Hour, M3hour, Mhour, Month)</td>
<td>1/day, 1/month, 5/month</td>
</tr>
<tr>
<td>12</td>
<td>MOA</td>
<td>2</td>
<td>14K</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CERESlib/Perl_Lib</td>
<td></td>
<td>131K</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>54</strong></td>
<td><strong>1,075K</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Recent Activity

22 Software & Data Deliveries Since May 5th, 2015

- **CATALYST (4)**
  - v2.0 Server, Ops Console & PGE Handlers
  - Intermediate Database updates
- **PR Tool (5)**
  - PR Tool v2.1 Web and DB deliveries
  - CATALYST v2.0 SSL support
  - Ed4 Variable display updates
- **Instrument (5)**
  - NPP Pre-Processor for EDOS Level-0 format
  - 2 scripting updates supporting EDOS NPP data flow testing
  - Postprocessor corrections for FM5
- **Clouds (1)**
  - Updates to Ed1 NPP PGEs
- **Inversion (2)**
  - Update Ed1 NPP PGE to run on dedicated P7 blades
- **TISA Gridding (2)**
  - Ed4 SSF1deg PCF corrections
  - Ancillary Xtk lookup tables
- **TISA Averaging (2)**
  - Ed3 SSF1deg redelivered to Use GMT and Ed4 format
  - Ed4 SSF1deg correction & x86 only
- **Perl_Lib (1)**
  - Support for new IBM P7 blades
<table>
<thead>
<tr>
<th>Product</th>
<th>Platform</th>
<th>Processed through</th>
<th>Publically Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS</td>
<td>Terra &amp; Aqua</td>
<td>May 31, 2015</td>
<td>Yes</td>
</tr>
<tr>
<td>SSF</td>
<td>Terra &amp; Aqua</td>
<td>May 31, 2015 (FM1 &amp; FM3 only)</td>
<td>Yes</td>
</tr>
<tr>
<td>SFC</td>
<td>Terra &amp; Aqua</td>
<td>April 2015</td>
<td>Yes</td>
</tr>
<tr>
<td>SSF1deg-Day/Month</td>
<td>Terra &amp; Aqua</td>
<td>June 2010</td>
<td>No*</td>
</tr>
<tr>
<td>SYN1deg</td>
<td>Merged</td>
<td>March 2015</td>
<td>Yes</td>
</tr>
<tr>
<td>ISCCP-D2like-Day/Nit</td>
<td>Terra &amp; Aqua</td>
<td>April 2015</td>
<td>Yes</td>
</tr>
<tr>
<td>ISCCP-D2like-GEO</td>
<td>GEO</td>
<td>April 2015</td>
<td>Yes</td>
</tr>
<tr>
<td>ISCCP-D2like-Mrg</td>
<td>Terra + Aqua + GEO</td>
<td>April 2015</td>
<td>Yes</td>
</tr>
<tr>
<td>SSF Edition 1-CV</td>
<td>Terra Aqua</td>
<td>July 31, 2015</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>July 16, 2015</td>
<td></td>
</tr>
<tr>
<td>Product</td>
<td>Platform</td>
<td>Processed Through</td>
<td>Expected Current</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>-------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>BDS</td>
<td>Terra &amp; Aqua</td>
<td>Dec 31, 2014</td>
<td>Nov 2015</td>
</tr>
<tr>
<td>SSF Beta2-Ed4</td>
<td>Terra Aqua</td>
<td>May 2, 2013 / May 10, 2013</td>
<td>Dec 2015</td>
</tr>
<tr>
<td>SSF</td>
<td>Terra Aqua</td>
<td>Dec 31, 2012 / Nov 30, 2012</td>
<td>Jan 2016</td>
</tr>
<tr>
<td>SSF1deg-Day/Month</td>
<td>Terra Aqua</td>
<td>Jan 2005 – Nov 2009</td>
<td>February 2016</td>
</tr>
<tr>
<td>SYN1deg</td>
<td>Merged</td>
<td>Jan – June 2008</td>
<td>June 2016</td>
</tr>
<tr>
<td>CldTypHist</td>
<td>Merged</td>
<td>Start January 2016</td>
<td>June 2016</td>
</tr>
</tbody>
</table>

**SYN1deg anticipated Release Date October 14th, 2015**

**CldTypeHist Pending Code Delivery**
## NPP Edition 1 Product Availability

<table>
<thead>
<tr>
<th>Product</th>
<th>Platform</th>
<th>Processed Through</th>
<th>Expected Current</th>
<th>Publically Available*</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS</td>
<td>S-NPP</td>
<td>March 31, 2014</td>
<td>March 2016</td>
<td>Yes</td>
</tr>
<tr>
<td>SSF</td>
<td>S-NPP</td>
<td>March 31, 2012*</td>
<td>April 2016</td>
<td>Yes</td>
</tr>
<tr>
<td>SSF1deg-Hour</td>
<td>S-NPP</td>
<td><strong>Start October 2015</strong></td>
<td>May 2016</td>
<td></td>
</tr>
<tr>
<td>SSF1deg-Day/Month</td>
<td>S-NPP</td>
<td><strong>Start October 2015</strong></td>
<td>May 2016</td>
<td></td>
</tr>
</tbody>
</table>

* SSF Ed1 NPP processing was stopped in Spring 2015 to correct a problem that limited the number of footprints and processing resumed in August 2015
Production Monitoring
Production Status Reporting

CERES Monthly Processing objectives and status to be reported on Confluence

• Monthly page created by DMT outlining what to process by stream
• ASDC to provide weekly status updates each Monday

Plan to begin keeping 6-month processing objectives page on Confluence to address longer-term objectives

https://asdcjira.larc.nasa.gov:9443/display/CARS/DMT+Processing+Objectives+-+August+2015
System Status
System Status

ASDC System Maintenance Window September 18th – 23rd
Start Friday 9/18 @ 5:00pm - End Wednesday 9/23 @ 9:00am

• Upgrade all GPFS clients
• Firmware updates to IBM and SGI disk controllers
• Coil-blue hardware refresh
• Migrate IDL license server from asdsun to coil-blue

New IBM Flex blades available in AMI-P – currently “dedicated” for Edition1 NPP SSF processing

• 6 IBM P7 Flex blades in production (2 in test area)
• 3 IBM x86 Flex blades in production (1 in test area)
• All Flex blades running Suse Linux 11.3 (vs 11.1)
• 20 additional x86 blades have been purchased for AMI-P
NPP-FM5 Data Flow Changes
Starting in September 2015, CERES FM5 data for NPP will no longer be obtained through NOAA IDPS

- EDOS to connect to JPSS ground station stream & package to EOS Level-0 format files
- All beta testing for CERES complete – Using small pre-processor PGE to merge 24-hour CERES file with 2-hour Attitude and Ephemeris files into 24-hour CERES files
- IDPS will no longer process RDRs

This approach anticipated to be used for FM6 on J1 and for RBI
Automation Framework

**CERES AuTomAted job Loading sYSTem (CATALYST)**

- Database system stores PRs and streamlines creation process
- Workflow manager server submits jobs to existing computing environment
- GUI enables operators to manage production at high level (exceptions)
- Framework exists as layer above existing production cluster
CATALYST Status

Release 1.0 Design proof-of-concept Clouds & Inversion only

CATALYST Release 1.0 began running Beta2-Ed4 Clouds and Inversion April 17\textsuperscript{th}, 2014

- 157 months Terra, 129 months Aqua
- >440,000 jobs successfully run

First build learning experience

- User feedback identified needed feature enhancements
- Bugs discovered

CATALYST Release 2.0 delivered for testing 8/24 and testing ongoing
38 Change Requests submitted based on version 1.0

• 35 directly implemented or addressed in release 2.0
• 1 identified as too difficult to implement
• 2 delayed, likely to be superseded by release 2.0 capabilities

Improvements by category

• PR Validation
• Problem Isolation
• PGE Management
• Operator’s Console Improvements
• Stability & Monitoring
• Logging
• Security
CATALYST 2.0 Improvements

PR Validation
• Extensive PR validation logic added – verify invalid or conflicting submitted PRs

Problem Isolation
Separate CATALYST to 6 sub-processes + Master process
• Limits client application downtime (i.e. PR Tool & Ops Console)
• Enables better logging and status reporting
• Improves Stability – Master processes monitors component health
• Contains known exception cases (i.e. server crashes, offline database, offline SGE)
CATALYST Build 2.0

**Kernel** – PR validation & coordination (PGE Handlers live here)

**Cluster** – Interfaces with SGE

**User** – Authenticates users against LDAP

**LogDB** – Interface to job record PostgreSQL

**ANGe** – Monitors ANGe archive ingest status

**XML-RPC** – API for Operator’s Console & PR Tool

**Master** – Minimal main process to manage all of the above
CATALYST 2.0 Improvements

PGE Management – file level

- Each PGE implemented as individual Perl module
- Modules installed in “CATALYST_HOME/handlers” directory
- PGE modules loaded on start of CATALYST kernel process
- Add/Remove PGE achieved by copy/delete .pm and restart kernel process
CATALYST 2.0 Improvements

PGE Management – Within Operator’s Console

- Whole Subsystem or individual PGEs can be enabled or disabled within UI.
- Epilogue settings can be adjusted by subsystem or PGE.
- PGE and Epilogue settings modifiable only by privileged CATALYST user (viewable by all).
- Epilogues run in FIFO order – if predecessor disabled follow-on will wait (prevents out of order file removal).
CATALYST 2.0 Improvements

Operator’s Console Improvements

- Added text search capability – search by date, job status, epilogue status or job ID
- Job status reporting
  - Added progress bars for jobs completed/failed
  - Color coded job listings (complete, in progress, failed, waiting)
- Progress bars and activity logs for user actions
- Read-Only Mode replaces need for v1.0 recovery utility – when server started, all job status’ viewable before administrator initializes jobs
- Inline documentation with hovertext
CATALYST 2.0 Improvements

This CATALYST job's science process has made an invalid memory reference.
CATALYST 2.0 Improvements

Operator’s Console Improvements Cont’d

• PRs organized by processing stream
• Console window is resizable (previously static)
• Intelligent button disabling based on permissions

Functions grayed out if insufficient user permissions
CATALYST 2.0 Improvements

Stability & Monitoring

• Added auto-detection for inactive jobs in SGE grid engine
  • Detects if job manually deleted ("qdel-ed") SGE jobs
  • Detects if blade issue causes job to hang
• New management utilities to monitor, start & stop CATALYST processes (Admin user only)
  • Graphical utility within Operator’s Console UI
  • Stand alone command line tool
• Improved system health monitoring email reporting
  • Generated by server process (previously standalone script)
  • More detailed and instructive info in reports – color coded to indicate status
CATALYST 2.0 Improvements

Logging
- Significantly improved logging capability
  - CATALYST Processes-level logging with all entries time stamped
  - Job details window now gives access within Console to:
    - PCF
    - PCF logs
    - LogStatus files
    - SGE logs
    - Epilogue logs

Security
- SSL support enabled for CATALYST API
  - All comm. From PR Tool & Operator’s Console now encrypted end-to-end
  - No visible changes to user
CATALYST Schedule

Schedule Re-baselined late April 2015 (on schedule since)

TRR – August 17th

V2.0 Delivery – August 25th

Associated PR Tool Delivery – August 25th

Release to ASDC – September 4th

ASDC Testing – September 7th – 25th

ORR – October 5th

Build 2.0 Promote – October 13th

 Deliver Ed4 SSF1deg & SYN1deg PGEs – October 23rd
Expanding CATALYST PGEs

PGE handlers to be delivered October 23rd 2015
- Edition 4 Inversion
- Edition 4 SSF1deg-Hr & Day/Month
- Edition 4 TSI, SARB & SYN1deg

PGE Handlers to be delivered December 2015
- Edition 1 NPP Clouds & Inversion

PGE Handlers to be delivered January 2016
- All Instrument and ERBE-Like PGEs
Backup Slides