CERES Data Management Group Status Report

CERES Science Team Meeting

May 5th, 2015

Jonathan Gleason

Data Management Team

Systems/ Optimization:
Nelson Hillyer
Michael Linsinbigler
Josh Wilkins

Production / Optimization:
Lisa Coleman
Carla Grune

Instrument:
Denise Cooper
Thomas Grepiotis
Dianne Snyder

Convolution:
Igor Antropov

Inversion:
Victor Sothcott

ERBElike:
Dale Walikainen
Jeremie Lande

SARB:
Tom Caldwell

Product Web Support:
Churngwei Chu
Ed Kizer
Cristian Mitrescu

Configuration Management:
Tammy Ayers
Joanne Saunders

Clouds:
Sunny Sun-Mack
Ricky Brown
Yan Chen
Liz Heckert
Rita Smith
Sharon Gibson

TISA Gridding:
Raja Raju

TISA Averaging:
Cathy Nguyen
Dennis Keyes

Task Manager:
Walt Miller

The Science Directorate at NASA's Langley Research Center

Overview

Current Data Availability

Recent DMT Activities (Deliveries)

CATALYST Status

Systems Status

Future Activities

The Science Directorate at NASA's Langley Research Center

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

Subsystem Number	Subsystem Name	Number of PGEs	LOC (to nearest 1K)	Publicly Available Data Products	Product Frequency
1	Instrument/Pre-Processor	2	9K		
1	Instrument	5 + lib	157K	BDS	1/day
2	ERBE-like/ Inversion	5 + lib	31K	ES-8	1/day
3	ERBE-like/ TSA	2	14K	ES-9, ES-4	1/month
4.1 – 4.4	Clouds/VIIRS Subset Code	1	21K		12/hour
4.1 – 4.4	Clouds	11	358K		
4.5 – 4.6	Inversion	11	110K	SSF	1/hour
5	SARB	1	36K		1/hour
6 & 9	TISA-Gridding	7	34K	SSF1deg-Hour, ISCCP-D2like-Day/Nit	60/month, 36/month, 1/month
11	GGEO	1	6K	ISCCP-D2like-GEO	1/month
7.2	Synoptic SARB	1	47K		
7.1 & 8 10	TISA-Averaging	4	102K	SSF1deg-Day, SSF1deg-Month, SYN1deg-(3Hour, M3hour, Mhour, Month)	1/day, 1/month, 1/month 5/month
12	MOA	2	14K		
	CERESlib/Perl_Lib		131K		
	Total	**53**	**1,070K**		

Recent Activity

22 Software & Data Deliveries Since October 6th, 2014

- Instrument (3)
  - 2 PGEs for new S_NPP EDOS data flow

- Clouds (3)
  - NPP updates
  - 2 script deliveries

- Inversion (3)
  - NPP updates
  - SOURCE TSI input file
  - Script updates – check all valid input

- SARB (2)
  - Ed 4 SYNI code
  - Memory request script updates

- TISA Gridding (1)
  - Ed4 SSFdeg-Hour code

- TISA Averaging (3)
  - Ed 4 SSF1deg-Day/Month, TSIB, & SYN1deg codes

- Perl_Lib (1)
  - Match DPO layout changes

- CERESlib (1)
  - Added support for Edition4 TISA products

- PR Tool (4)
  - Web Interface 2.0
  - Add 6 new PGEs (Ed4 & EDOS changes)
  - 2 Routine updates

- GGEO (1)
  - 1 Data-only delta delivery

The Science Directorate at NASA's Langley Research Center

Edition 3 Product Availability

Product	Platform	Processed through	Publically Available
BDS	Terra & Aqua	December 31, 2014	Yes
SSF	Terra & Aqua	November 31, 2014 (FM1 & FM3 only)	Yes
SFC	Terra & Aqua	October 2014	Yes
SSF1deg-Day/Month	Terra & Aqua	March 2014	No*
SYN1deg	Merged	October 2014	Yes
ISCCP-D2like-Day/Nit	Terra Aqua	November 2014 November 2014	Yes
ISCCP-D2like-GEO	GEO	November 2014	Yes
ISCCP-D2like-Mrg	Terra + Aqua + GEO	November 2014	Yes
SSF Edition 1-CV	Terra Aqua	March 31, 2015	No

Edition 4 Product Availability

Product	Platform	Processed Through	Expected Current	Publically Available*
BDS	Terra Aqua	December 31st, 2013	October 2015	Yes
SSF Beta2-Ed4	Terra Aqua	November 18th, 2011 May 28th, 2011	December 2015	No
SSF	Terra Aqua	December 31st, 2009	January 2016	Yes
SSF1deg-Hour	Terra Aqua	Jan – June 2005	February 2016	June 10th, 2015
SSF1deg-Day/ Month	Terra Aqua	**Start July 2015**	February 2016	Aug 12th, 2015
SYN1deg	Merged	**Start May 2015**	February 2016	July 8th, 2015
CldTypHist	Merged	**Start January 2016**	March 2016	March 9th, 2016

* Public availability dates assume release after first 2 years processed

NPP Edition1 Product Availability

Product	Platform	Processed Through	Expected Current	Publically Available*
BDS	S-NPP	March 31, 2014	March 2016	Yes
SSF	S-NPP	May 31, 2012*	April 2016	Yes
SSF1deg-Hour	S-NPP	**Start August 2015**	May 2016	October 2015
SSF1deg-Day/Month	S-NPP	**Start August 2015**	June 2016	October 2015

* SSF Ed1 NPP to be reprocessed starting June 2015 to resolve missing footprint issue

Production Automation Effort

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

The Science Directorate at NASA's Langley Research Center

CATALYST Status

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

CATALYST Release 1.0 began running Beta2-Ed4 Clouds and Inversion April 17th, 2014
- 139 months Terra, 104 months Aqua
- >380,300 jobs successfully run

First build learning experience
- User feedback identified needed feature enhancements
- Bugs discovered

CATALYST What We've Learned

- Users need more summary & "at-a-glance" info in UI

- Need strong search function in UI: Search by date range, by exit codes, or by job status

- Should handle ASDC wrapper scripts same as other PGE jobs

- Log Database should be updatable on-the-fly

- Err on the side of inclusion when it comes to log files

- Must account for external job failures (e.g. manual qdel)

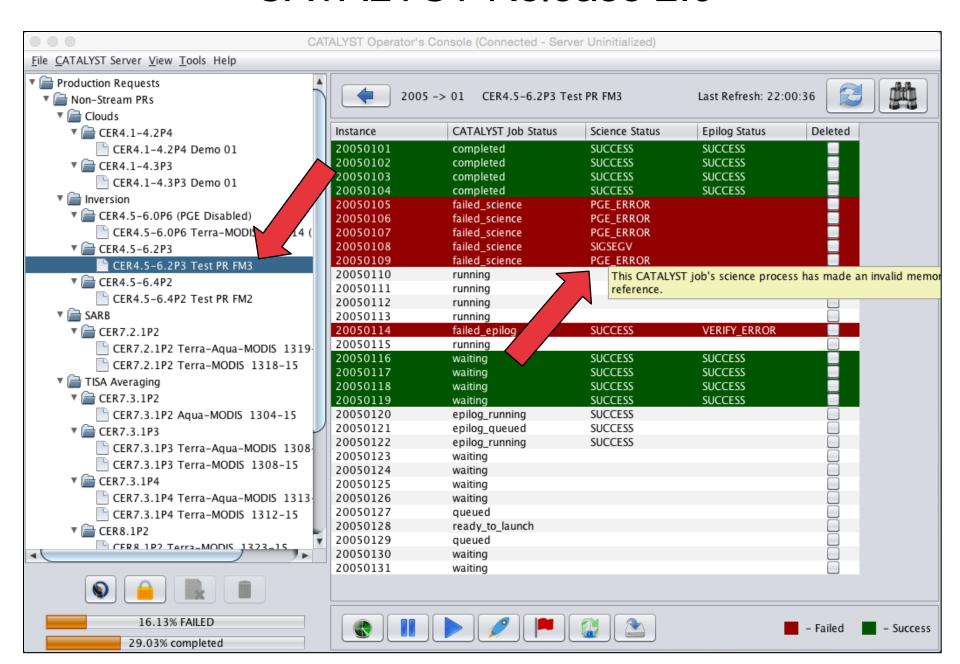- As anticipated several bugs discovered and corrected

CATALYST 2.0 Improvements

User Interface changes/enhancements

- Color coding and progress bars to show job summary info
- Added search function: by date range or exit status
- Dynamically sizable windows in console
- Epilog wrapper scripts available in Console – one stop shop
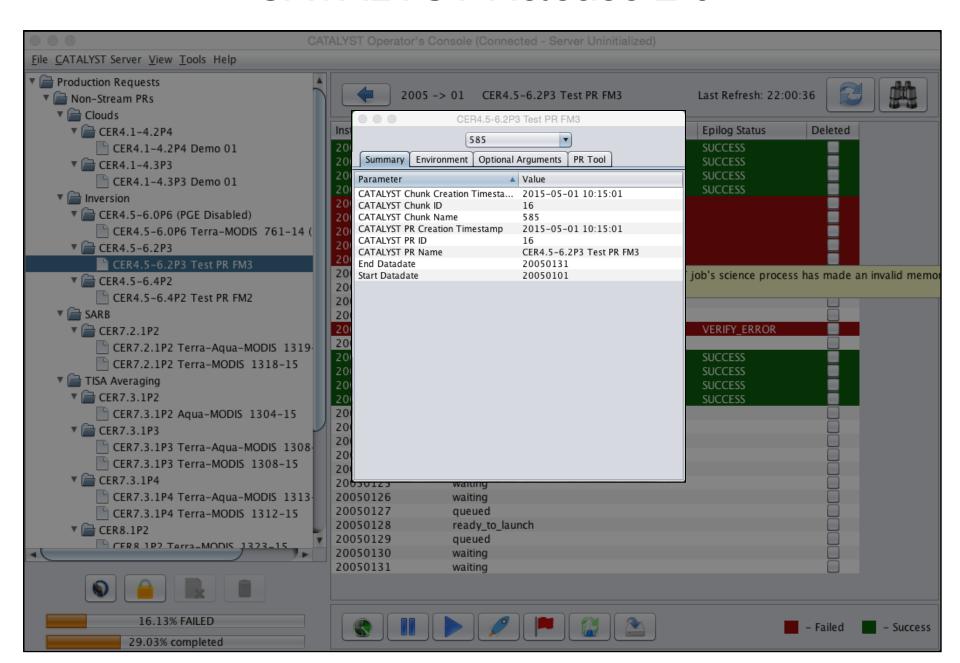- PGEs hanlders can be enabled/disabled dynamically

Server-side significant changes

- Server thread management restructured for maintainability
- Internal data stricture improvements – expedited load & search
- Actively track jobs in grid engine – no longer passive
- Implemented API to keep log DB dynamically updated
- Never check PR input – range check all PRs
- Expanded logging

CATALYST Release 2.0

CATALYST Release 2.0

CATALYST Build 2.0 Schedule

5/5/15

Key Milestones	FY15		
	2Q'15	3Q'15	4Q'15
CATALYST B2.0 Final Development	1/12 ▬▬▬▬▬▬▬▬ 5/28		
B2.0 DMT Unit Testing		5/29 ▬ 6/4	
B2.0 Test Readiness Review		△ 6/4	
CATALYST B2.0 Delivery		△ 6/5	
B2.0 CM Testing		6/8 ▬ 6/17	
B2.0 Release to ASDC		△ 6/17	
B2.0 SI&T Testing		6/18 ▬ 7/8	
B2.0 Operational Readiness Review			△ 7/23
Promote CATALYST B2.0			△ 7/24
Deliver Monitoring Utility			△ 7/24
Deliver Ed4 PGE Handlers			△ 7/31

▬▬▬ Work Effort   △ Open Milestone   ▲ Completed Milestone

Expanding CATALYST PGEs

PGE handlers to be delivered July 2015

- Edition 4 Inversion
- Edition 4 SSF1deg-Hr & Day/Month
- Edition 4 TSI, SARB & SYN1deg

PGE Handlers to be delivered September 2015

- Edition 1 NPP Clouds & Inversion

PGE Handlers to be delivered October 2015

- All Instrument and ERBE-Like PGEs

PR Tool 2.0

PR Tool 1.0 live since March 2013

Release 2.0 promoted January 2015 – implemented user feedback
- Search function improvements
  - Added free-text search: search by date range, by PR # range
  - Multivariate search criteria specification
- Dynamically filters search page options based on menu selections
- Updates to PR approval and state change capabilities

Two interim deliveries to add new Edition 4 PGEs and new FM5-EDOS pre-Processor PGE

Release 2.1 – Create and Search PRs by Processing Stream
- Search by processing stream implemented and under review
- Create all PRs for Stream in single interface (vs. per PGE)

Planned Software Deliveries and Data Product Availabilities

The Science Directorate at NASA's Langley Research Center

System Migration Update

As of March 2015 all CERES production code runs on AMI-P system (P6 and x86 chips)

- Magneto system failed during March 2nd power outage
- Previously Clouds and Inversion Ed1-CV ran on Magneto

New IBM P7 blades available within AMI-P and complemented by 2 new x86 blades, all running Suse Linux 11.3

- Current AMI-P production blade OS Suse Linux 11.1
- P7 blades will support Clouds production

DMT testing for all Subsystems, all PGEs planned in May & June 2015

- Subsystem representatives identified to conduct recompile and non-recompile test cases for SLES 11.3 for all PGEs
- P6 PGEs will also test on IBM P7 blades

New 11.3 blades to support NPP Clouds and Inversion processing in near term

DMT, SCCRs & JIRA

JIRA is issue & project tracking software used widely at ASDC and in industry

- Implements customized processes

- Can enable dashboard-like reporting

- Can link to SVN and wiki tools via Confluence software

CATALYST and PR Tool teams have had success using JIRA for issue and software delivery tracking

DMT will use JIRA to augment PGE software delivery process

- Streamline issue communication between DMT and ASDC

- Investigation for how to implement pending:

  - Store SCCR as JIRA ticket vs augment existing process with JIRA ticket

  - Use existing SCCR DB and requirements log DB or replace

Backup Slides

MODIS C5 vs C6

MODIS Collection 5 soon to be replaced by Collection 6

Disk space considerations

- C5 & C6 ~7TB/Year
- Combined 371+TB not practical to keep on spinning disk

Planned Approach

- Move all C6 onto disk (~182TB)
- Keep only C5 seasonal months on disk (~59TB)

Result: 130TB disk space savings (241TB vs 371TB combined total)

Product	Science Delivery	Target Public Release
Edition 1 NPP BDS & SSF (Time varying Gains & SRFs)	Delivered	October 6th, 2014
Ed4 Inversion	Delivered	October 30th, 2014
Ed4 SSF-deg-Day/Month	October 17th, 2014	January 15th, 2015
Ed4 TSI code	December 5th, 2104	March 30, 2015
Ed4 SYNI code	December 12th, 2014	
Ed4 SYN1deg	December 19th, 2014	

Planned Delivery and Release Milestones (2 of 2)

Product	Science Delivery	Target Public Release
Ed3 SSF1deg-Month	November 21st, 2014	January 30th, 2015
Ed4 ISSCP-D2like-Day/Nit	January 30th, 2015	April 15th, 2015
Ed4 ISCCP-D2like-GEO	February 13th, 2015	April 30th, 2015
Ed4 ISCCP-D2like-Mrg	February 27th, 2015	May 15th, 2015
Ed4 Flux-By-Cloud-Type	March 27th, 2015	June 15th, 2015