OUTLINE

Team
Highlights
Code Re-architecture
Systems
DATA MANAGEMENT TEAM

SSAI Management:
Walter Miller
Susan Thomas

Denise Cooper
Thomas Grepiotis
Hunter Winecoff
Dianne Snyder
Dale Walikainen

Sunny Sun-Mack
Ricky Brown
Steve Kohler
Yan Chen
Elizabeth Heckert
Rita Smith
Walt Miller

Tom Caldwell

Victor Sothcott
Igor Antropov

PC Sawaengphokhai
Jay Garg
Hunter Winecoff

Nelson Hillyer
Tammy Ayers
Dennis Keyes

Dale Walikainen
Jeremie Lande

Josh Wilkins
Cathy Nguyen
Ed Kizer
Beau Branch

Carla Grune
Liz Heckert
HIGHLIGHTS
Team Activities & Code Deliveries
Ed1B NOAA-20
&
Ed2A S-NPP

Data products released April 1, 2022
DEVELOPMENTS

• Code supporting finalization of new editions:
  • Ed2A S-NPP
  • Ed1B NOAA-20

• CATALYST server automatic restart

• Prompt PHP security upgrades
  • CERES PR Tool
  • CERES Website

• Applying Deep Blue aerosols over land to S-NPP SSFs
• Snow & ice data loss mitigation:
  • 557th Weather Wing – supplementary data
  • snow & ice map, “ice age” value

• Production support for upcoming EBAF 4.2 Surface

• Daily albedo coverage fix for FluxByCldTyp
DEVELOPMENTS

- MOA: MERRA2 and GEOS-IT accommodation
- New ASDC ingest system, Dark Horse: Forward and backwards compatibility
- PIV-M compatibility
- FLASHFlux:
  - 3 out of 7 PGEs running in CATALYST
  - MOA, Clouds in production; Inversion delivered
## ED 4 PRODUCT AVAILABILITY

<table>
<thead>
<tr>
<th>Product</th>
<th>Platform</th>
<th>Processed Thru</th>
<th>Publicly Available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSF1deg-Hour</td>
<td>Terra, Aqua</td>
<td>Jan. ‘22</td>
<td>Yes</td>
</tr>
<tr>
<td>SSF1deg-Day/-Month</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYN1deg-1Hour/MHour</td>
<td>Terra+Aqua</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYN1deg-Day/-Month</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## ED 4 PRODUCT AVAILABILITY

<table>
<thead>
<tr>
<th>Product</th>
<th>Platform</th>
<th>Processed Thru</th>
<th>Publicly Available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CldTypHist</td>
<td>Terra+Aqua</td>
<td>Dec. ‘21</td>
<td>Yes</td>
</tr>
<tr>
<td>FluxByCldTyp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBAF</td>
<td></td>
<td>Nov. ‘21</td>
<td></td>
</tr>
<tr>
<td>EBAF ToA</td>
<td></td>
<td>Jan. ‘22</td>
<td></td>
</tr>
</tbody>
</table>
## S-NPP PRODUCT AVAILABILITY

<table>
<thead>
<tr>
<th>Edition</th>
<th>Product</th>
<th>Platform</th>
<th>Processed Thru</th>
<th>Publicly Available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ed2A</td>
<td>BDS</td>
<td>S-NPP</td>
<td>Jan. '22</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SSF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSF1deg-Hour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSF1deg-Day/-Month</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ed1A</td>
<td>SYN1deg-Hour/MHour</td>
<td>Terra+S-NPP</td>
<td>Nov. '17*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SYN1deg-Day/-Month</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* L3 processing stopped. Instrument in RAPS mode.
## ED 1B NOAA-20 PRODUCT AVAILABILITY

<table>
<thead>
<tr>
<th>Product</th>
<th>Platform</th>
<th>Processed Thru</th>
<th>Publicly Available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS</td>
<td>NOAA-20</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>SSF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSF1deg-Hour</td>
<td>NOAA-20</td>
<td>Jan. ‘22</td>
<td></td>
</tr>
<tr>
<td>SSF1deg-Day/-Month</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>
LIBERA DATA MGMT WORKING GROUP

• Ad hoc meetings: Libera and Radiation Budget Science Project (RBSP)

• Libera team discussed high-level plan:
  • Science Data Center design
  • Science Processing plan

• Working Libera-RBSP Interface Control Document (ICD) soon
CODE RE-ARCHITECTURE
TISA & Clouds
TISA CODE RE-ARCHITECTURE

Current Codebase

- Bloated
- Redundancies
- Disjointed
  - Scattered input data sources
  - Tailored routines per source
## TISA CODE RE-ARCHITECTURE

<table>
<thead>
<tr>
<th>Current Codebase</th>
<th>New Codebase</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Bloated</td>
<td>• Conducive to scientists’ needs</td>
</tr>
<tr>
<td>• Redundancies</td>
<td>• Experimentation</td>
</tr>
<tr>
<td>• Disjointed</td>
<td>• Extension</td>
</tr>
<tr>
<td>• Scattered input data sources</td>
<td>• Similar data stored together</td>
</tr>
<tr>
<td>• Tailored routines per source</td>
<td>• Standardized netCDF API</td>
</tr>
<tr>
<td></td>
<td>• Common library functions to reuse</td>
</tr>
<tr>
<td></td>
<td>• Driven out cyclomatic complexity</td>
</tr>
<tr>
<td></td>
<td>(nested statements)</td>
</tr>
</tbody>
</table>
TISA CODE RE-ARCHITECTURE

Object-Oriented Gridding C++

Gridding containers, zones, regions

JSON property configuration files

Property containers

Object-Oriented "like" Interpolation & Averaging Fortran
TISA CODE RE-ARCHITECTURE

Current Framework

New Framework

1 source file, 2 subroutines
### TISA CODE RE-ARCHITECTURE

<table>
<thead>
<tr>
<th>Software Lines of Code</th>
<th>Original Code</th>
<th>Refactored Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>nb2bb_lwwn_ind_day</td>
<td>255</td>
<td>nb2bb_lwwn_general</td>
</tr>
<tr>
<td>nb2bb_lwwn_indNit</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>nb2bb_lwwn_ocn_day</td>
<td>258</td>
<td></td>
</tr>
<tr>
<td>nb2bb_lwwn_ocnNit</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>nb2bb_lwwn_sno_day</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td>nb2bb_lwwn_snoNit</td>
<td>245</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1522</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Difference: 1366 lines**

**90% code reduction!**
CLOUDS CODE RE-ARCHITECTURE

CLOUD MASK CODE REFACTORIZING
CLOUDS CODE RE-ARCHITECTURE

• Focus on CERESmask – cloud mask

• Powerful code, decision tree
  • cloud, no cloud
  • other classification types
CLOUDS CODE RE-ARCHITECTURE

• Maintainability challenges:
  • 7,121 lines of code
  • Direct memory access
  • Inputs generated within codebase

• New features with refactoring:
  • Object-oriented programming
    • Parent/child inheritance of methods
    • Less code duplication
    • Enables unit testing
    • Polymorphism: one interface used on multiple data types

• Configuration files — easier to tune repeated values, don’t touch source code
Current listing of variables and functions

<table>
<thead>
<tr>
<th>CERESmask</th>
<th>cont'd</th>
<th>cont'd</th>
<th>cont'd</th>
<th>cont'd</th>
</tr>
</thead>
<tbody>
<tr>
<td>CitlTab</td>
<td></td>
<td></td>
<td></td>
<td>C5_Test()</td>
</tr>
<tr>
<td>phiRange</td>
<td></td>
<td></td>
<td></td>
<td>C6_Test()</td>
</tr>
<tr>
<td>T4Range</td>
<td></td>
<td></td>
<td></td>
<td>C1_AVHRRTtest()</td>
</tr>
<tr>
<td>threshGT1</td>
<td></td>
<td></td>
<td></td>
<td>C2_AVHRRTtest()</td>
</tr>
<tr>
<td>T37_oldCalib</td>
<td></td>
<td></td>
<td></td>
<td>C3_AVHRRTtest()</td>
</tr>
<tr>
<td>T11_oldCalib</td>
<td></td>
<td></td>
<td></td>
<td>C4_AVHRRTtest()</td>
</tr>
<tr>
<td>T12_oldCalib</td>
<td></td>
<td></td>
<td></td>
<td>C5_AVHRRTtest()</td>
</tr>
<tr>
<td>delt37</td>
<td></td>
<td></td>
<td></td>
<td>C6_AVHRRTtest()</td>
</tr>
<tr>
<td>delt12</td>
<td></td>
<td></td>
<td></td>
<td>C2_GOES12Test()</td>
</tr>
<tr>
<td>Terra37_origV4</td>
<td></td>
<td></td>
<td></td>
<td>C3_GOES12Test()</td>
</tr>
<tr>
<td>sigma_3</td>
<td></td>
<td></td>
<td></td>
<td>C4_GOES12Test()</td>
</tr>
<tr>
<td>thresh23_glint</td>
<td></td>
<td></td>
<td></td>
<td>C5_GOES12Test()</td>
</tr>
<tr>
<td>polat</td>
<td></td>
<td></td>
<td></td>
<td>C6_GOES12Test()</td>
</tr>
<tr>
<td>numPhi</td>
<td></td>
<td></td>
<td></td>
<td>ref213_016</td>
</tr>
<tr>
<td>numT4</td>
<td></td>
<td></td>
<td></td>
<td>cserf213_016</td>
</tr>
<tr>
<td>TerraMODISData</td>
<td></td>
<td></td>
<td></td>
<td>cserfSTD213_016</td>
</tr>
<tr>
<td>AquaMODISData</td>
<td></td>
<td></td>
<td></td>
<td>cserf213_show</td>
</tr>
<tr>
<td>MODISData</td>
<td></td>
<td></td>
<td></td>
<td>cserf213 snd</td>
</tr>
<tr>
<td>VIRSData</td>
<td></td>
<td></td>
<td></td>
<td>cserf213_snd</td>
</tr>
</tbody>
</table>
Class structure – refactored mask code

- **Chunk**
  - m_nLinePixelCategories
  - m_nChunkResponseCategories
  - m_vLine0PixelCategories
  - m_nLineDataTypes

- **Data Type**
  - AVHRR
  - VIRS
  - MODIS
  - MODIS05
  - MODISExtend
  - AODIS
  - AODIS05
  - MAC
  - GOES
  - GOES_6
  - GOES_10
  - GOES_12
  - GOES_13
  - GOES_14
  - MSG_SVIRI
  - HIASAVARAHI
  - GOES_R_18_ABI
  - GOES_R_17_ABI
  - MSG
  - MSG_41
  - NTSAT_45
  - NTSAT_50

- **Pixel**
  - m_fLatitude
  - m_fLongitude
  - m_fPixelZenithAngle
  - m_iChannelSource
  - m_bPolar
  - m_iLightCondition
  - m_iPosition
  - m_iDataTypes

- **LightCondition**
  - DAY
  - TWILIGHT
  - NIGHT

- **CloudMask**
  - m_bPolar
  - m_f_latitude
  - m_f_longitude
  - m_iChannelSource
  - m_iLightCondition
  - m_iPixelCategoryValues

- **CloudMaskNighttime**
  - determineClearMaskCategory()
  - determineCloudMaskCategory()

- **CloudMaskTwilight**
  - determineClearMaskCategory()
  - determineCloudMaskCategory()

- **CloudMaskTime**
  - determineClearMaskCategory()
  - determineCloudMaskCategory()

- **Pixel**
  - processPixel()

- **PixelGoes**
  - processPixel()

- **PixelModisAqua**
  - processPixel()

- **PixelModisTerra**
  - processPixel()

- **PixelMag**
  - processPixel()

- **PixelMsat**
  - processPixel()

- **PixelMirs**
  - processPixel()

- **Singleton LUT_CERESMaskConfig** replaces magic numbers embedded in code base
CLOUDS CODE RE-ARCHITECTURE

SERVERLESS ARCHITECTURE EXPERIMENT
CLOUDS CODE RE-ARCHITECTURE

• **Goal**: Run CURRENT CERES Clouds code in AWS

• **Motivations**:
  • Refactoring (new code) – long timeline
  • 1 week of processing / 1 year of data
  • Currently working CERES Edition 5 development – iterate

• **Build capacity** on CERES for running
  • In Docker
  • On AWS
  • With “serverless” architecture – lambdas
AWS lambda: compute service to run code WITHOUT provisioning or managing servers.
CLOUDS CODE RE-ARCHITECTURE

• Previous work: “containerizing” Clouds PGEs
  • Docker images for 2 product generation executives (PGEs):
    • Clouds main – “1P6”
    • Clouds clear radiance history (CRH) – “2P5”
  • Enabled:
    • Building Clouds binaries
    • System agnostic execution

• Current work: building & running Clouds Docker images in AWS
CLOUDS CODE RE-ARCHITECTURE

Developer Workstation

CodeCommit

CodeBuild

CodeBuild Instance

Pull Source Code

Build Executables

Build Docker Images

Deploy Docker Images

Update Lambdas

Elastic Container Registry

Lambda Functions

Compute Environment

Build Step

AWS Service

run unit tests
Current design:

• Employ lambda functions for creating PCFs and running PGEs

• Functional: launch-lambda for one PGE: 1P6
  (can also launch via command line)

• STDOUT text going to CloudWatch log streams

• Output data stored in S3 or EFS
CLOUDS CODE RE-ARCHITECTURE

Simple Storage Service

Elastic Filesystem

Command with run parameters

Launch 1P6

1P6 Hourly Retrieval Instance

1P6 Hourly Footprint Instance

1P6 Hourly Retrieval Instance

1P6 Hourly Footprint Instance

1P6 Hourly Retrieval Instance

1P6 Hourly Footprint Instance

1P6 Hourly Retrieval Instance

1P6 Hourly Footprint Instance

1P6 Hourly Retrieval Instance

1P6 Hourly Footprint Instance

Launch 2P5...

Lambda Instance

Data Storage Service


**CLOUDS CODE RE-ARCHITECTURE**

**Serverless Architecture**

**Advantages**
- No infrastructure maintenance
- Pricing based on compute time
- Encourages modular architecture
- Automated build/test/deploy
- 1P6 and 2P5 instances can run in parallel - 1,000 simultaneous lambdas!

**Caveats**
- Increased vendor lock-in – AWS-specific
- Steeper learning curve
- Not viable for every computing scenario
- 15 min. maximum runtime
- 10 GB memory/512 MB local storage limit
- Filesystem read-only (other than /tmp directory)
SYSTEMS

Ordering Tool & Hardware
HARDWARE

• First HPE w/ Qumulo purchase
  • 5 nodes in 2020
  • 1 PB usable space

• Second HPE w/ Qumulo purchase
  • 5 nodes in late Summer 2021
  • 1 PB usable space

• Additional 2 node purchase
  • Allows more usable space
  • Restriping

• Updated total usable space:
  ~3.2 PB
CERES ORDERING TOOL

- Develops and maintains custom, CERES visualization subsetting and ordering tool
- Internal and public versions
- Public ordering tool on virtualized hardware
CERES ORDERING TOOL

**NOV 2021**
Successfully tested subsetting in new physical location - Computational Research Facility (CRF)

**DEC 2021**
Released new version of CCCM (D1)
Quickly mitigated log4j vulnerability

**FEB 2022**
Transitioned to PIV-M authentication

**APR 2022**
Released NOAA-20 Ed1B and NPP Ed2A for ordering and subsetting

**UP NEXT**
Converting QA tool to virtual machines
SUMMARY

• Ed1B NOAA-20 and Ed2A NPP editions released

• Building capacity to run in AWS
  • Facilitates Edition 5 development
  • Anticipates “SIPS in the Cloud”

• New hardware boosts project storage

• Actively working with Libera Data Management Working Group