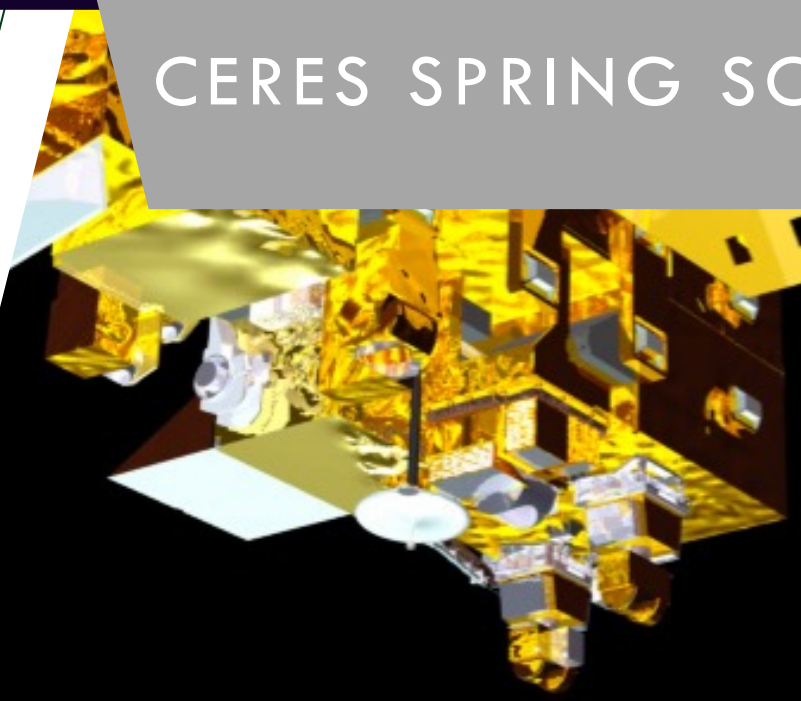
CERES DATA MANAGEMENT TEAM WORKING GROUP REPORT

CERES SPRING SCIENCE TEAM MEETING

Kathleen Dejwakh
April 26, 2022



OUTLINE

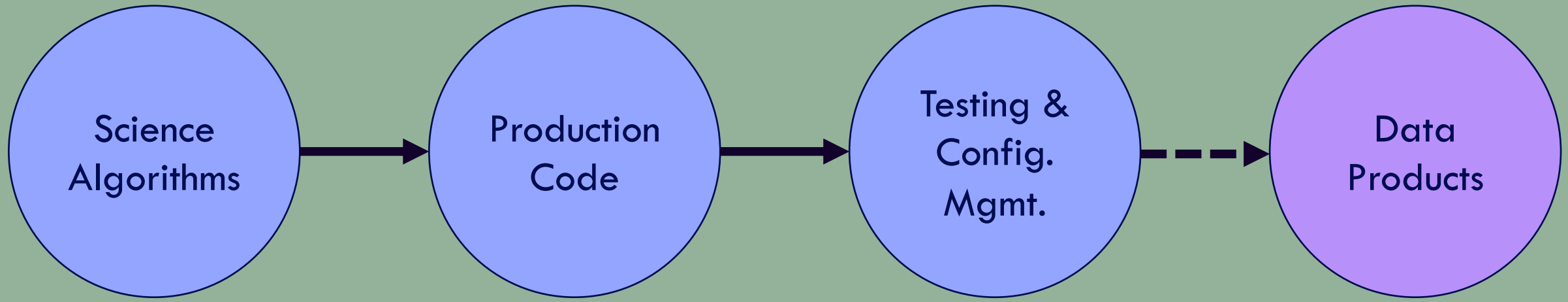
Team

Highlights

Code Re-architecture

Systems

DATA MANAGEMENT TEAM



DATA MANAGEMENT TEAM

SSAI Management:

Walter Miller
Susan Thomas

HIGHLIGHTS

Team Activities & Code Deliveries

Ed1B NOAA-20

&

Ed2A S-NPP

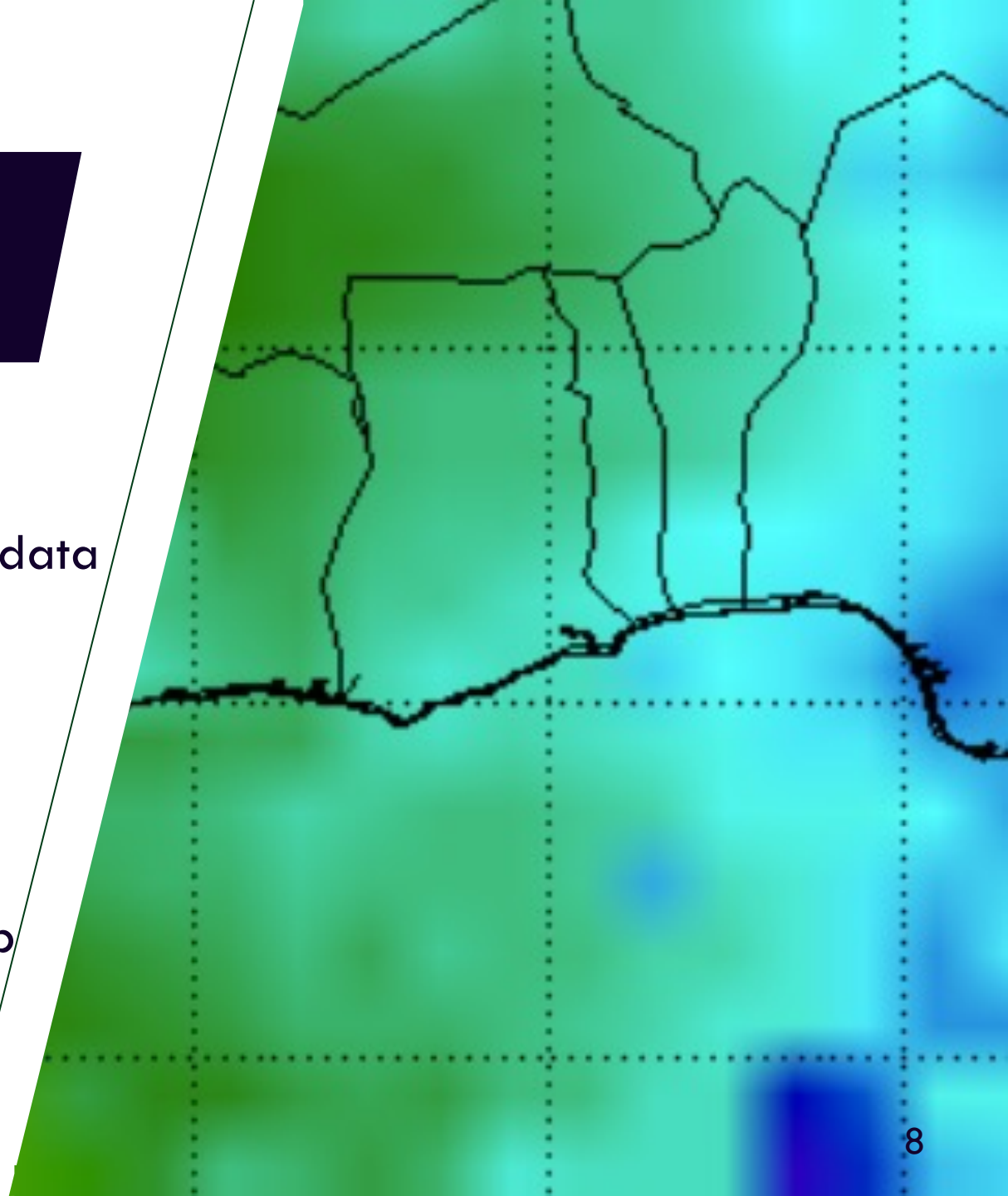
Data products released April 1, 2022

DELIVERIES

- 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

DELIVERIES

- 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



DELIVERIES

- 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

Product	Platform	Processed Thru	Publicly Available?
BDS	Terra, Aqua	Jan. '22	Yes
SSF			
SSF1deg-Hour			
SSF1deg-Day/-Month			
SYN1deg-1Hour/MHour	Terra+Aqua		
SYN1deg-Day/-Month			

ED 4 PRODUCT AVAILABILITY

Product	Platform	Processed Thru	Publicly Available?
CldTypHist	Terra+Aqua	Dec. '21	Yes
FluxByCldTyp			
EBAF		Nov. '21	
EBAF ToA		Jan. '22	

S-NPP PRODUCT AVAILABILITY

Edition	Product	Platform	Processed Thru	Publicly Available?
Ed2A	BDS	S-NPP	Jan. '22	Yes
	SSF			
	SSF1deg-Hour		Sept. '19*	
	SSF1deg-Day/-Month			
Ed1A	SYN1deg-Hour/MHour	Terra+S-NPP	Nov. '17*	
	SYN1deg-Day/-Month			

* L3 processing stopped. Instrument in RAPS mode.

ED 1B NOAA-20 PRODUCT AVAILABILITY

Product	Platform	Processed Thru	Publicly Available?
BDS	NOAA-20	Jan. '22	Yes
SSF			
SSF1deg-Hour			
SSF1deg-Day/-Month			

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

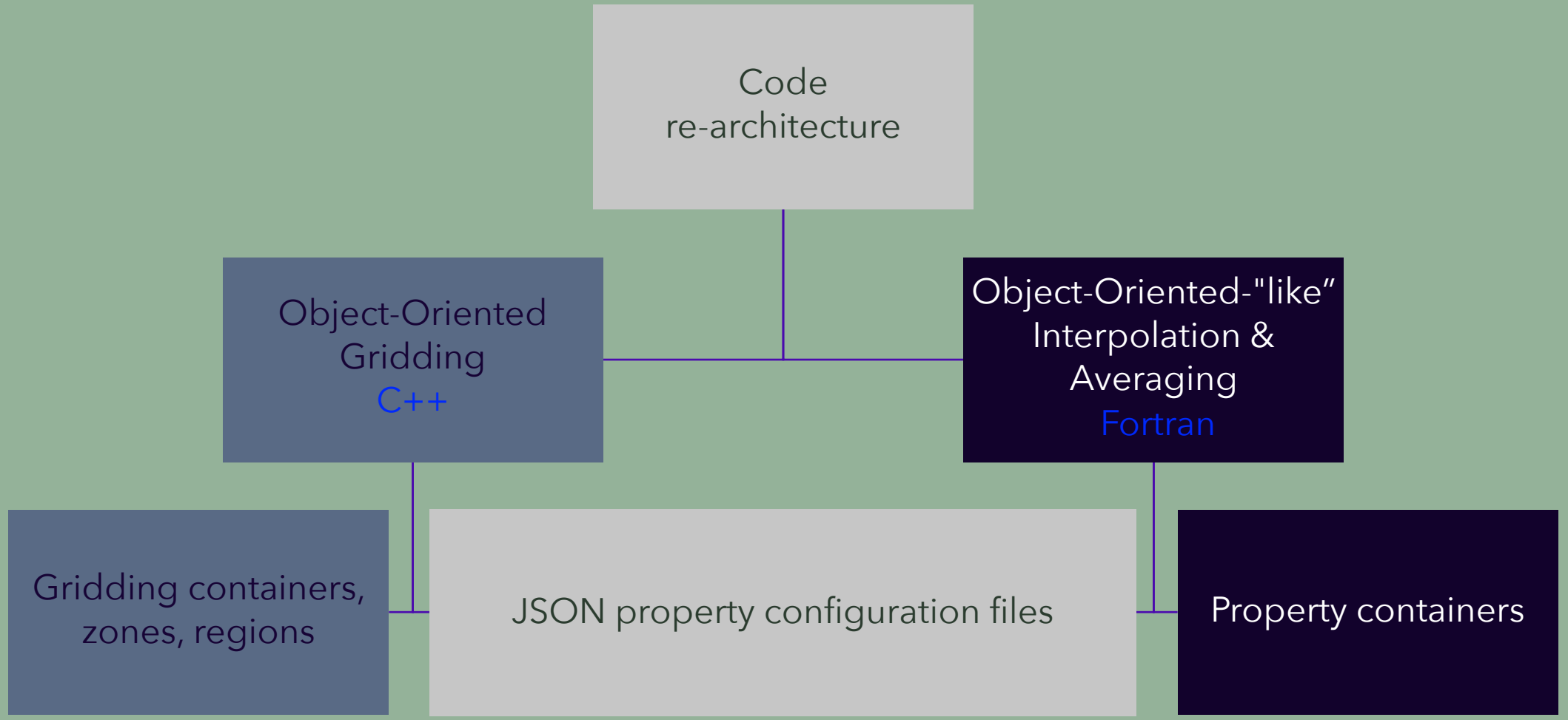
- Bloated
- Redundancies
- Disjointed
  - Scattered input data sources
  - Tailored routines per source

TISA CODE RE-ARCHITECTURE

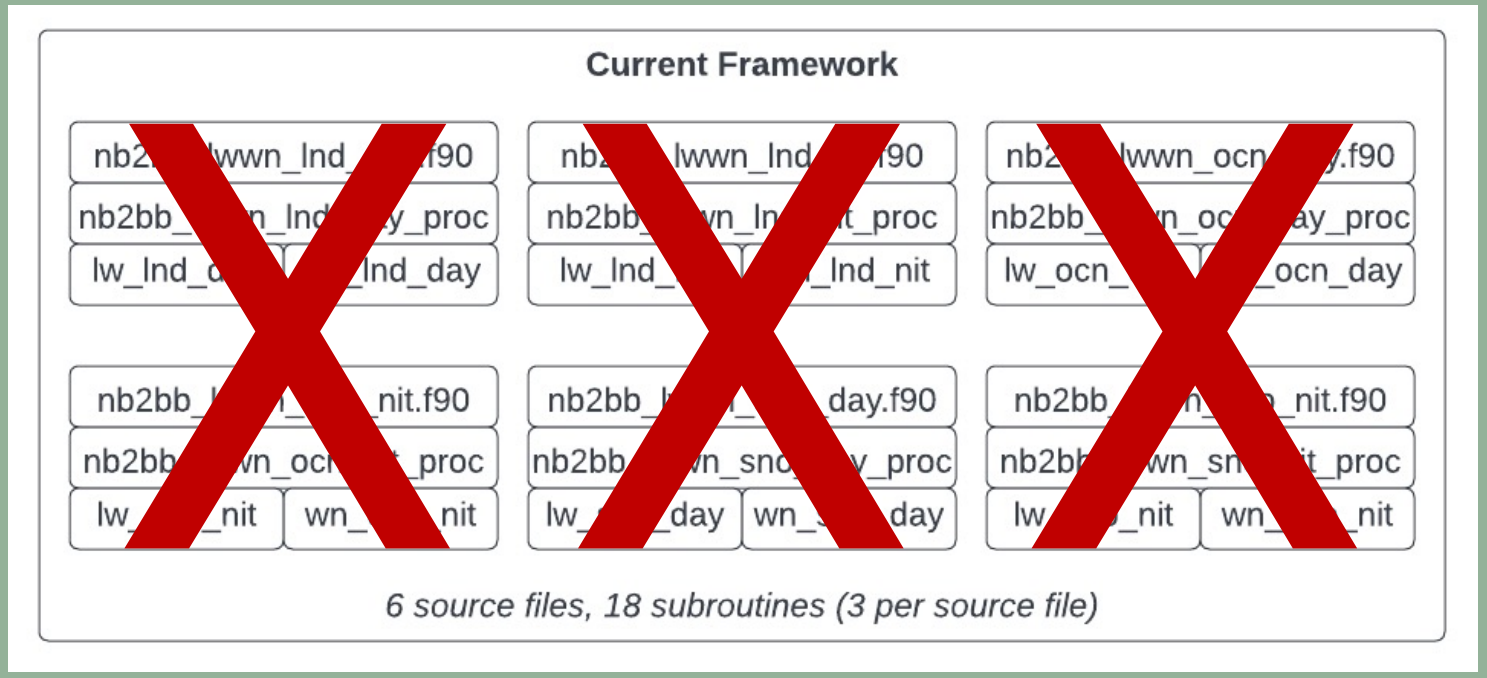
Current Codebase	New Codebase

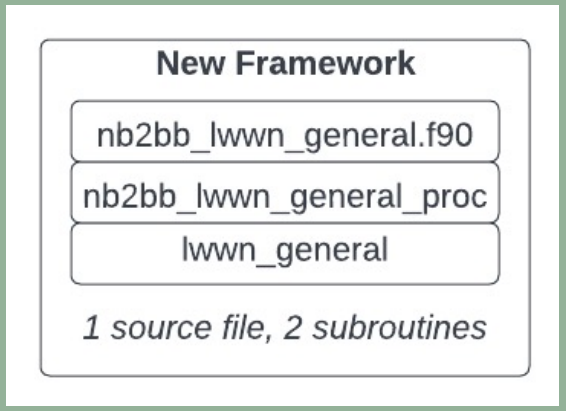
• Bloated
• Redundancies
• Disjointed
  • Scattered input data sources
  • Tailored routines per source
	• Conducive to scientists' needs
	  • Experimentation
	  • Extension
	• Similar data stored together
	• Standardized netCDF API
	• Common library functions to reuse
	• Driven out cyclomatic complexity (nested statements)

TISA CODE RE-ARCHITECTURE



TISA CODE RE-ARCHITECTURE





TISA CODE RE-ARCHITECTURE

Software Lines of Code			
Original Code		Refactored Code	
nb2bb_lwwn_lnd_day	255	nb2bb_lwwn_general	156
nb2bb_lwwn_lnd_nit	253		
nb2bb_lwwn_ocn_day	258		
nb2bb_lwwn_ocn_nit	255		
nb2bb_lwwn_sno_day	256		
nb2bb_lwwn_sno_nit	245		
Total	1522		
	Difference: 1366 lines		

90% code reduction!

CLOUDS CODE RE-ARCHITECTURE

CLOUD MASK CODE REFACTORING

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

CLOUDS CODE RE-ARCHITECTURE

Current listing of variables and functions

CERESmask

- CithrTab
- phiRange
- T4Range
- thrCi
- diff34_Cloud_Dust_thrsh
- elevat_limit
- SIRlow_STD
- ratio21
- ratio23
- ratio13
- ratio08_06
- ratio47_21
- ratio124_06
- diff34
- diff35
- diffcs34
- diffcs_4
- diffcs45
- diff065_cs
- diff67_11
- diff85_11
- diff85_67
- diff67_85
- diff45
- diff11_13
- ref213_016
- csref213_016
- csrefSTD213_016
- csref213_snow
- csref213STD_snow

cont'd

- T37_oldCalib
- T11_oldCalib
- T12_oldCalib
- delt37
- delt11
- delt12
- Terra37_origV4
- sigma_3
- thresh23_glint
- polat
- numPhi
- numT4
- TerraMODISData
- AquaMODISData
- MODISData
- VIRSData
- GOESData
- GOES12Data
- GOES8Data
- AVHRRData
- MSGData
- MTSATData
- FY2CData

cont'd

- coastFlag
- wrapandmask
- newPolFlag_inMask
- title
- printFlag
- calibGain0160
- calibOffset0160
- pixelstruct
- pixel
- waterCld_mean_tauLess1
- waterCld_STD_tauLess1
- iceCld_mean_tauGreat6
- iceCld_STD_tauGreat6
- sunglint_prob
- waterC
- sunGlint_glint_ratio31
- sunGlint_glint_ref1_1
- sunGlint_glint_btemp3
- sunGlint_glint_ref1_2
- night_water_btemp34_1
- night_water_btemp4
- night_water_btemp34_2
- night_water_btempcs4

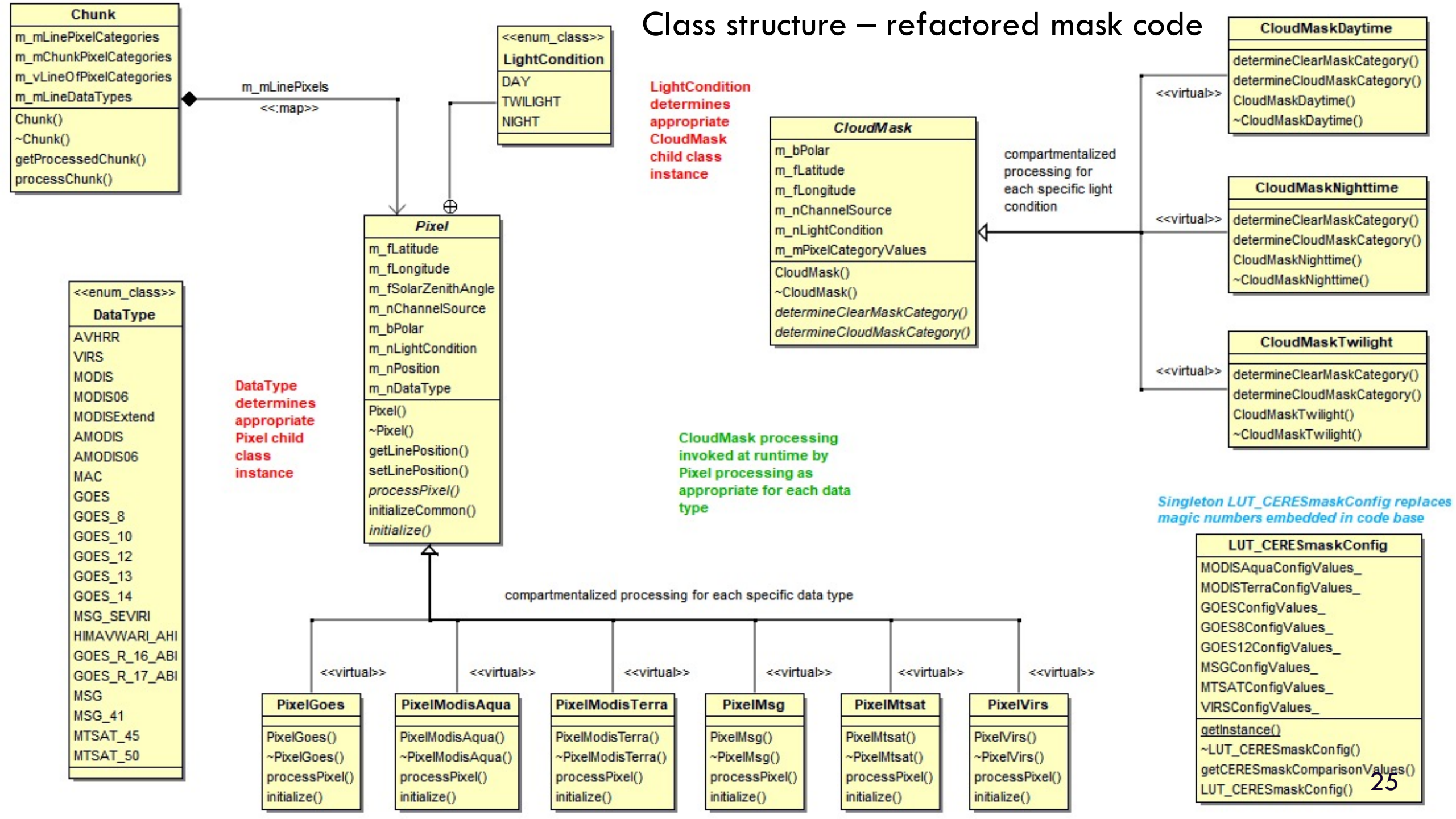
cont'd

- getceresfiremask_()
- getceresshadowmask_()
- getceresaerosolmask_()
- getceresaerotypemask_()
- getaeroacceptmask_()
- GetSERCAACiThre()
- DayTimeTests()
- DayTimeTerraPolarTests_thresh()
- DayTimeTerraPolarTests()
- DayTimeAquaPolarTests()
- DayTimeAVHRRTests()
- TwilightTests()
- NightTimeTests()
- NightTimeMODISPolarTests()
- TwilightMODISPolarTests()
- tlimTest()
- IRThresholdTest()
- visThresholdTest()
- vis16ThresholdTest()
- SR_IRThresholdTest()
- SR_IR_highThresholdTest()
- SR_IR_lowThresholdTest()
- ocean_aerosolTest()

cont'd

- C5_Test()
- C6_Test()
- C1_AVHRRTest()
- C2_AVHRRTest()
- C3_AVHRRTest()
- C4_AVHRRTest()
- C5_AVHRRTest()
- C6_AVHRRTest()
- C2_GOES12Test()
- C3_GOES12Test()
- C4_GOES12Test()
- C5_GOES12Test()
- E1_Test()
- E2_Test()
- E3_Test()
- E4_Test()
- E5_Test()
- E1_GOES12Test()
- E2_GOES12Test()
- E3_GOES12Test()
- nightimeTest()

Class structure – refactored mask code



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

lean

“

AWS lambda: compute service to run code WITHOUT provisioning or managing servers

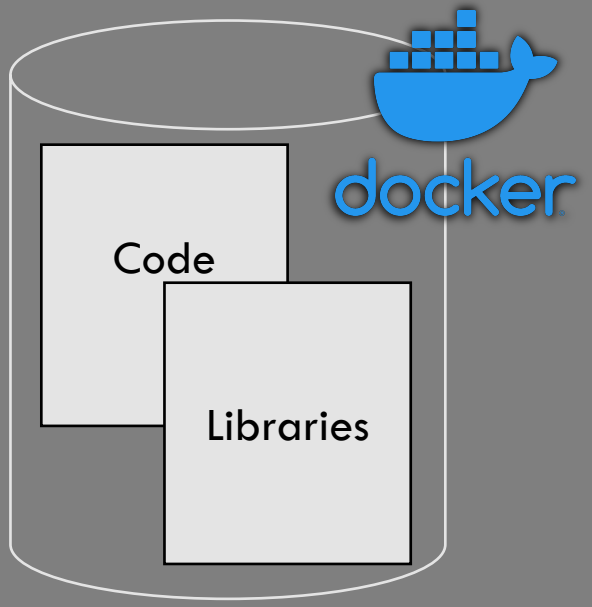
ephemeral

docs.aws.amazon.com ”

scalable

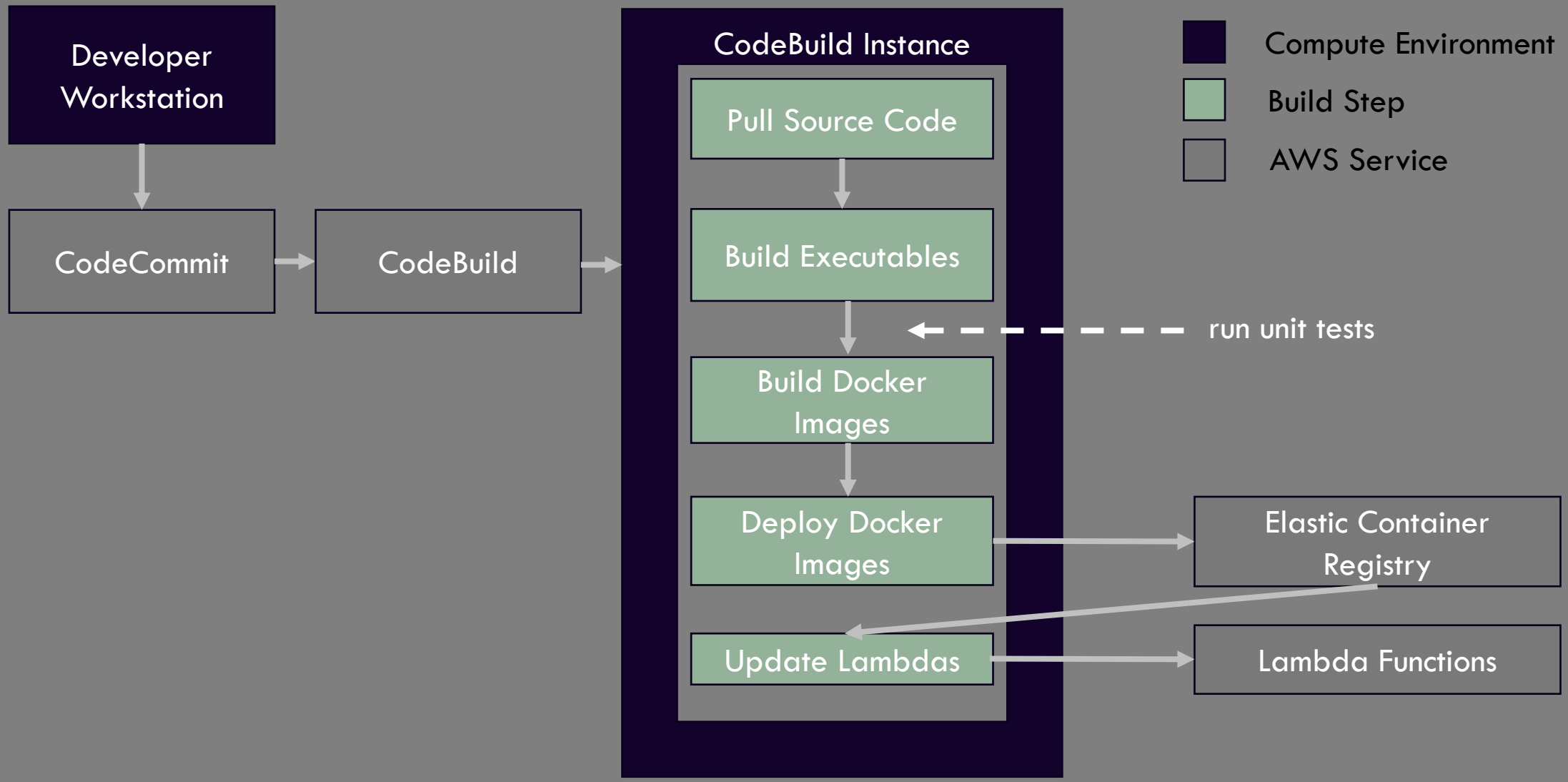
CLOUDS CODE RE-ARCHITECTURE

- 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

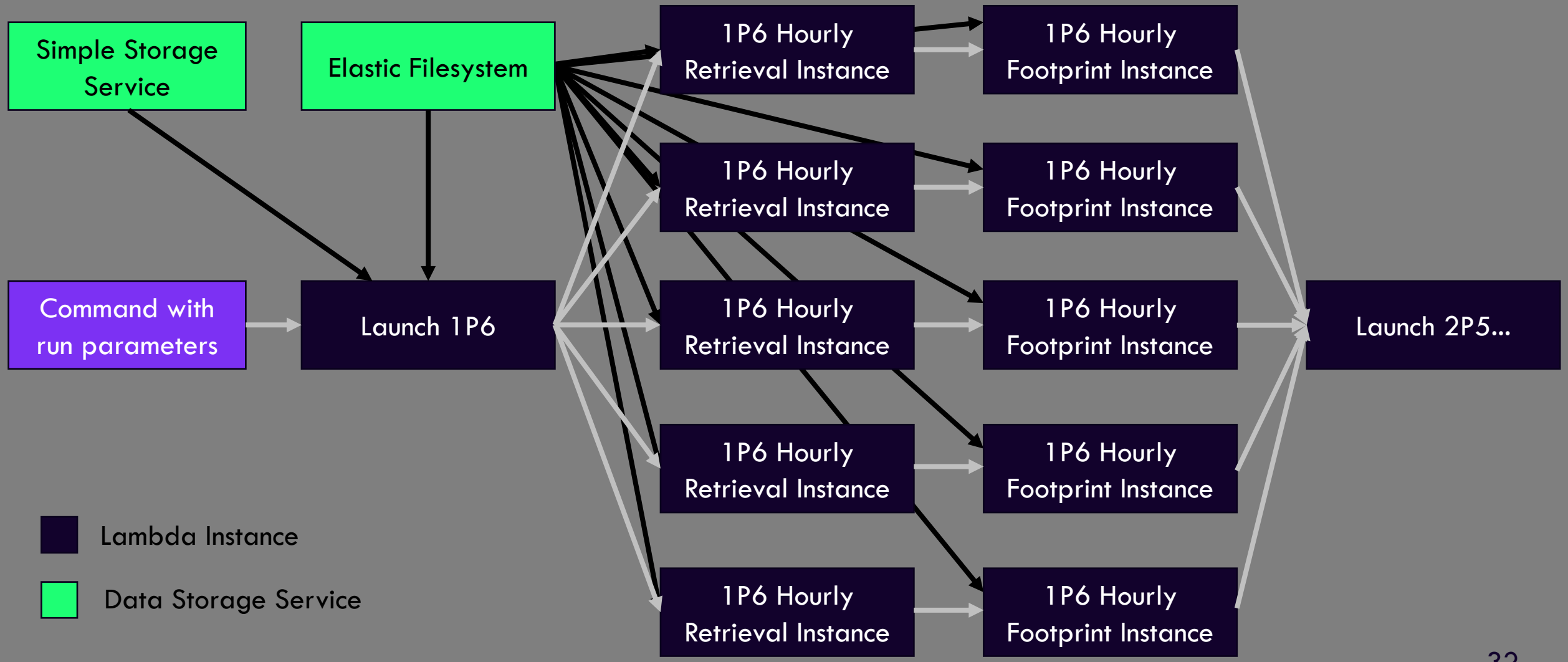


CLOUDS CODE RE-ARCHITECTURE

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



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

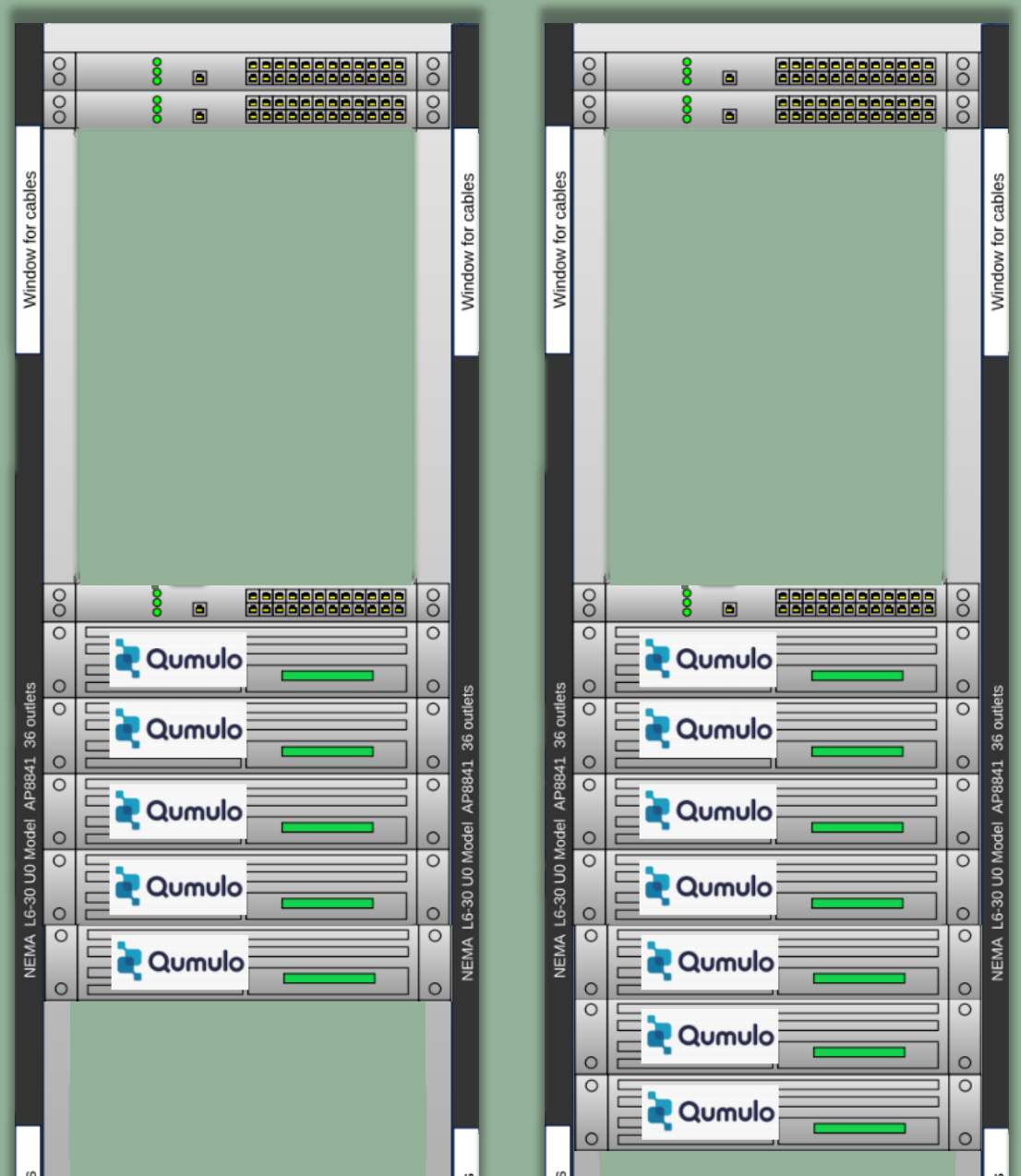
- Increased vendor lock-in – AWS-specific
- Steeper learning curve
- Not viable for every computing scenario
- 15 min. maximum runtime
- 10 GB memory/512 MB local storage limit
- Filesystem read-only (other than /tmp directory)

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

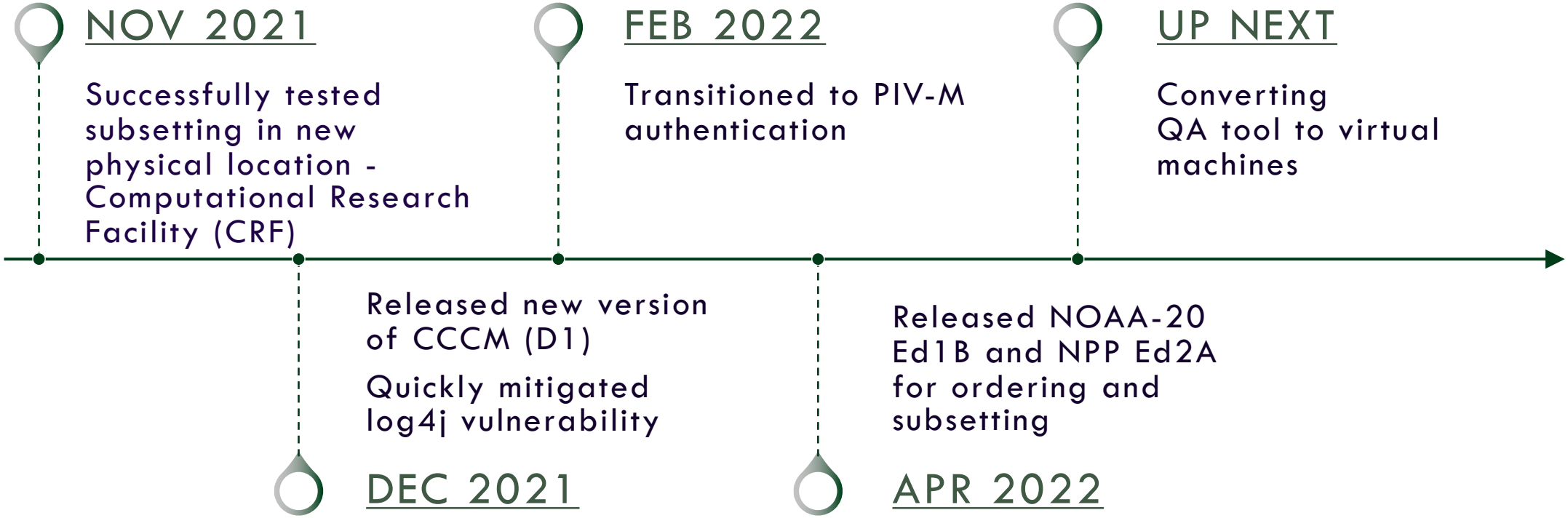


CERES ORDERING TOOL

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

Pamela Mlynczak
Churngwei Chu
Babak Samani

CERES ORDERING TOOL



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

Image credit: Mauricio Soto