

**Jet Propulsion Laboratory**  
California Institute of Technology

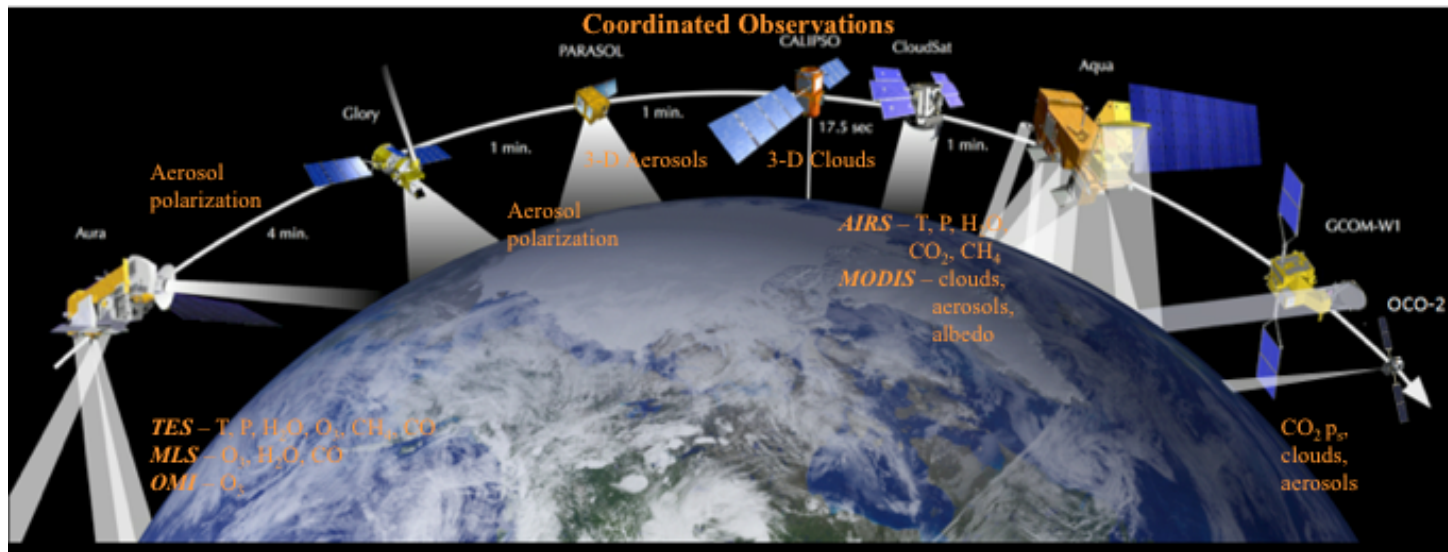
# Cloud On-boarding for OCO-2 and Sentinel- 1A/B

Thursday, July 27, 2017  
ESIP Federation 2017 Summer Meeting  
Bloomington, IN

**HySDS Team:** Hook Hua, Gerald Manipon, Michael Starch,  
Lan Dang, Justin Linick, Namrata Malarout  
NASA Jet Propulsion Laboratory / California Institute of Technology

Copyright 2017, by the California Institute of Technology. ALL RIGHTS RESERVED. United States Government Sponsorship acknowledged. Any commercial use must be negotiated with the Office of Technology Transfer at the California Institute of Technology

# Onboarding OCO-2 onto Cloud

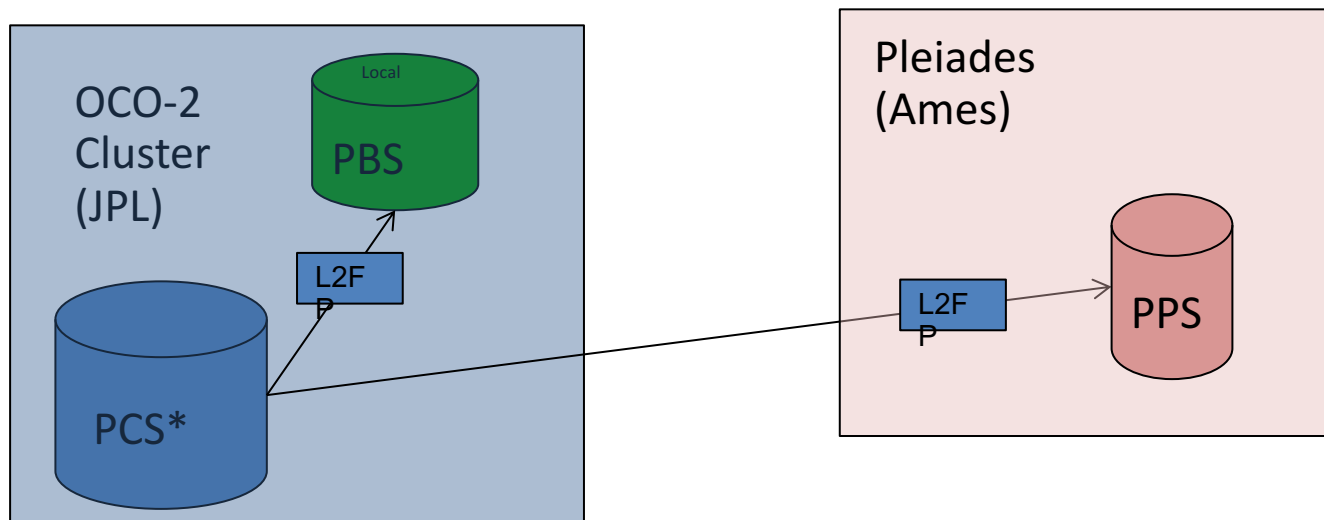


- OCO-2 launched on July 2nd, 2014, at the head of the A-Train
- Collect global measurements of atmospheric carbon dioxide with the precision, resolution, and coverage needed to characterize sources and sinks in order to improve our understanding of the global carbon cycle

# OCO-2 Science Data System



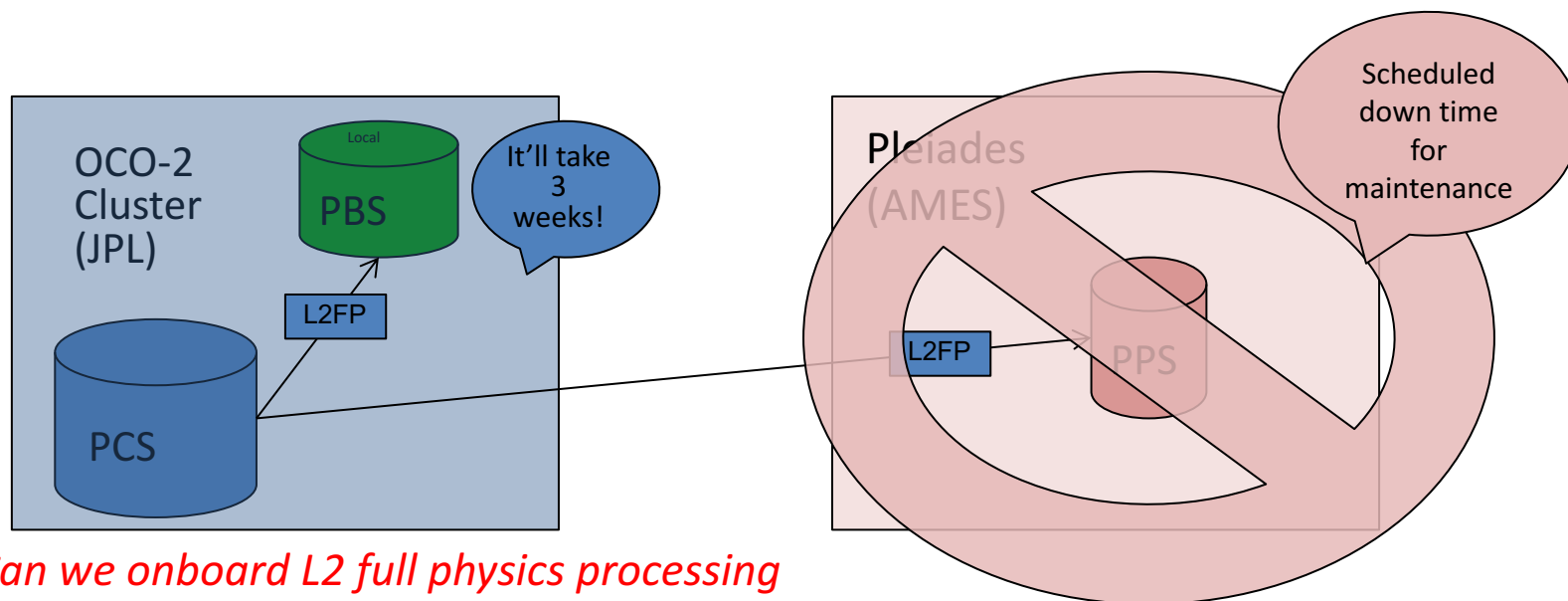
- NASA OCO-2 Science Data Operations System (SDOS)
  - Forward (1X) and bulk processing (4X)
  - L2 bulk processing ported to NASA AMES Pleiades Supercomputer
  - L2 full physics processing of granule soundings on ~200 nodes (15X)
  - Running 48 x PGE processors on each compute node



# Day 0

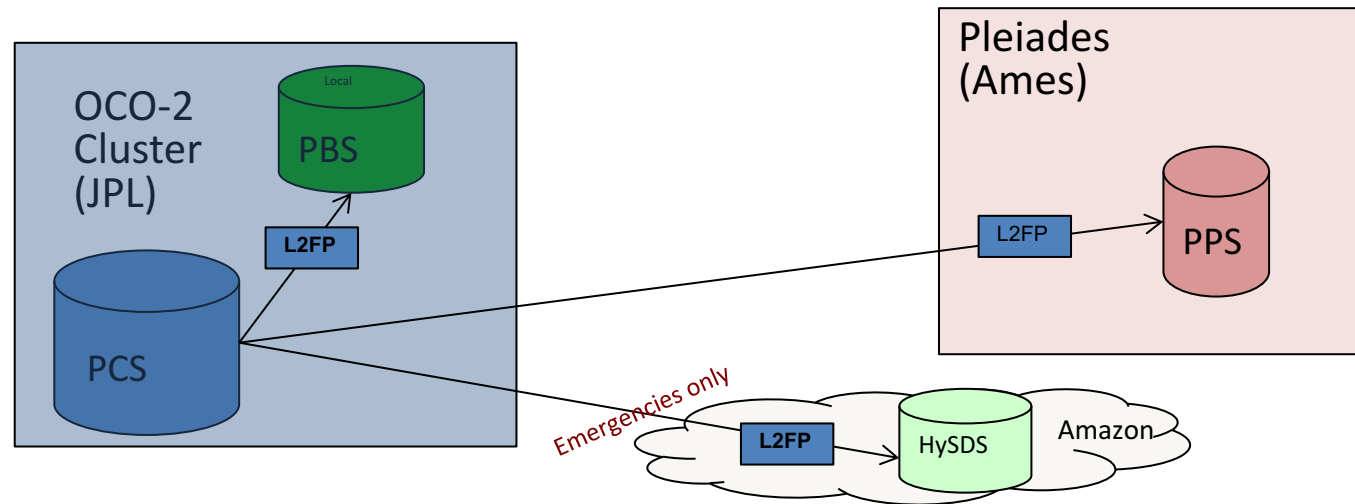


- Request: Process 3-months of data in 7 days (rate of 13x)
  - ~6% of data are usable soundings
  - OCO-2 cluster built to a 4x throughput requirement.
  - *BTW: Pleiades down for facility-wide maintenance*



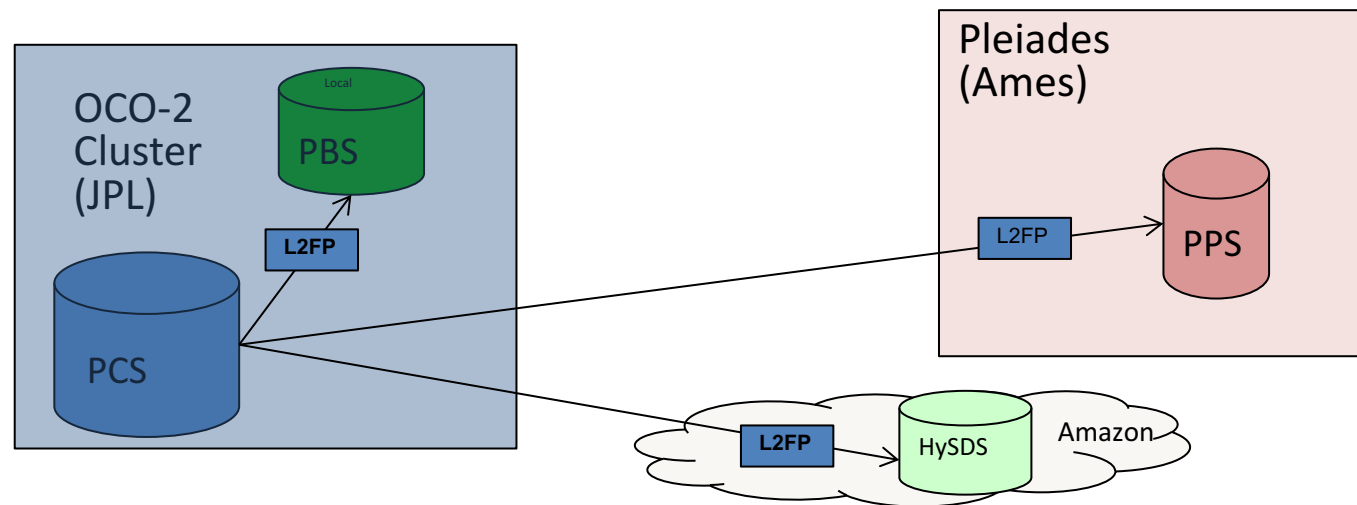
*Can we onboard L2 full physics processing onto AWS quickly?*

- Team successfully integrated L2FP executable into HySDS and demonstrated run on Amazon.
- L2FP developers validated outputs
- Plan larger dataset for end-to-end system test, to be vetted by science





- **Benchmarking** on Amazon to determine best and most cost-effective machine types
- **End-to-end** testing with SDOS.
- **Requirements changed!**
  - Reprocess ALL processable cloud-free soundings (~6% -> ~15% of data) for the past 9 months.
  - With data delivery to NASA GESDISC DAAC.
- Cloud computing approach now considered a **necessary** part of OCO-2 SDOS computing capabilities.



- **Benchmark analysis**
- **Production planning and estimates**
  - Compute, storage, egress
- **Migrated to AWS spot market**
  - Spot terminations..

equivalent	# soundings	processing time (# days)	total wall- clock hours	# compute instances needed	# jobs	US-West-2 (Oregon)	
						on-demand	
1 data-day	60,000	1.0	719.8	30	1875	\$775.94	
1 data-month	1,800,000	30.0	21593.8	30	56250	\$23,278.06	

*Cost estimated based on published rack rates of m2.4xlarge at the time*

- HySDS official **delivery and operational** handoff to OCO-2 SDOS.
- OCO-2 SDS in AWS **integrated** with JPL on-premise
- ***High-resiliency operations*** on AWS “spot market”
  - Up to 90% cheaper than “on-demand”
  - Fault tolerant across compute instance terminations
    - Spot terminations
    - Availability Zone (AZ) rebalancing



# Example Production Run



- **40X real-time processing**
  - 465 granules for October 2014—*processed in under a day*
- Data volumes
  - 1.6TB data products generated
  - 12.5TB data fed into processing pipeline
- Science Data System
  - EC2 in US-West-2 (Oregon)
- Storage
  - S3 in US-West-1 (Northern California)
- Compute
  - EC2 in US-US-West-2 (Oregon)
- Auto-scaling
  - 1000 x cc2.8xlarge / US-West-2 (Oregon)
  - **32,000 x l2\_fp simultaneous processors**

# What Did We Learn?

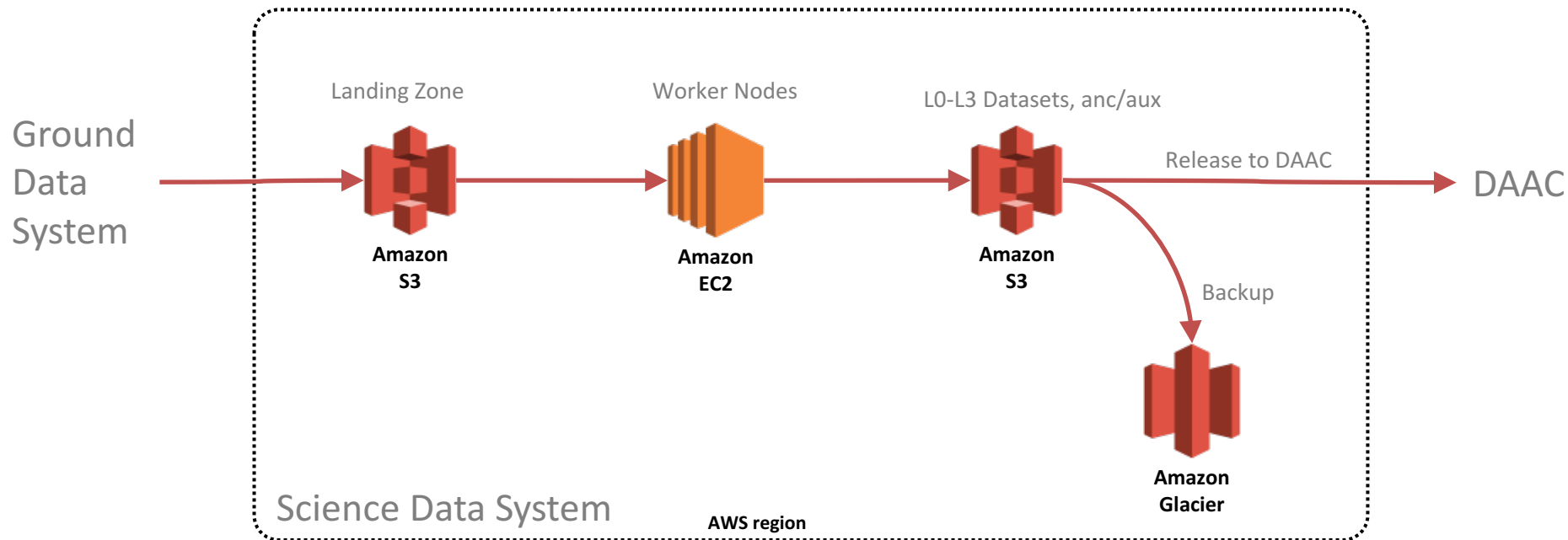


- Motivations
  - Frequent science **requirement changes**
  - Needed **agile** science data system approach
  - Increase in science computing needs
  - HPC Supercomputer scheduled **downtimes** may conflict with science processing requirements
  - Need **elastic and large-scale** processing capability
- Lessons Learned
  - More than just “*fork & lift*” into the cloud
  - Affordable if you can leverage **spot market** pricing
  - **Benchmarking and metrics** are key to good decision-making
  - **Large scaling affects**: Heavy operational use uncovers issues with robustness and operability for most common use cases

# Basic Premise of Cloud-based Science Data Processing



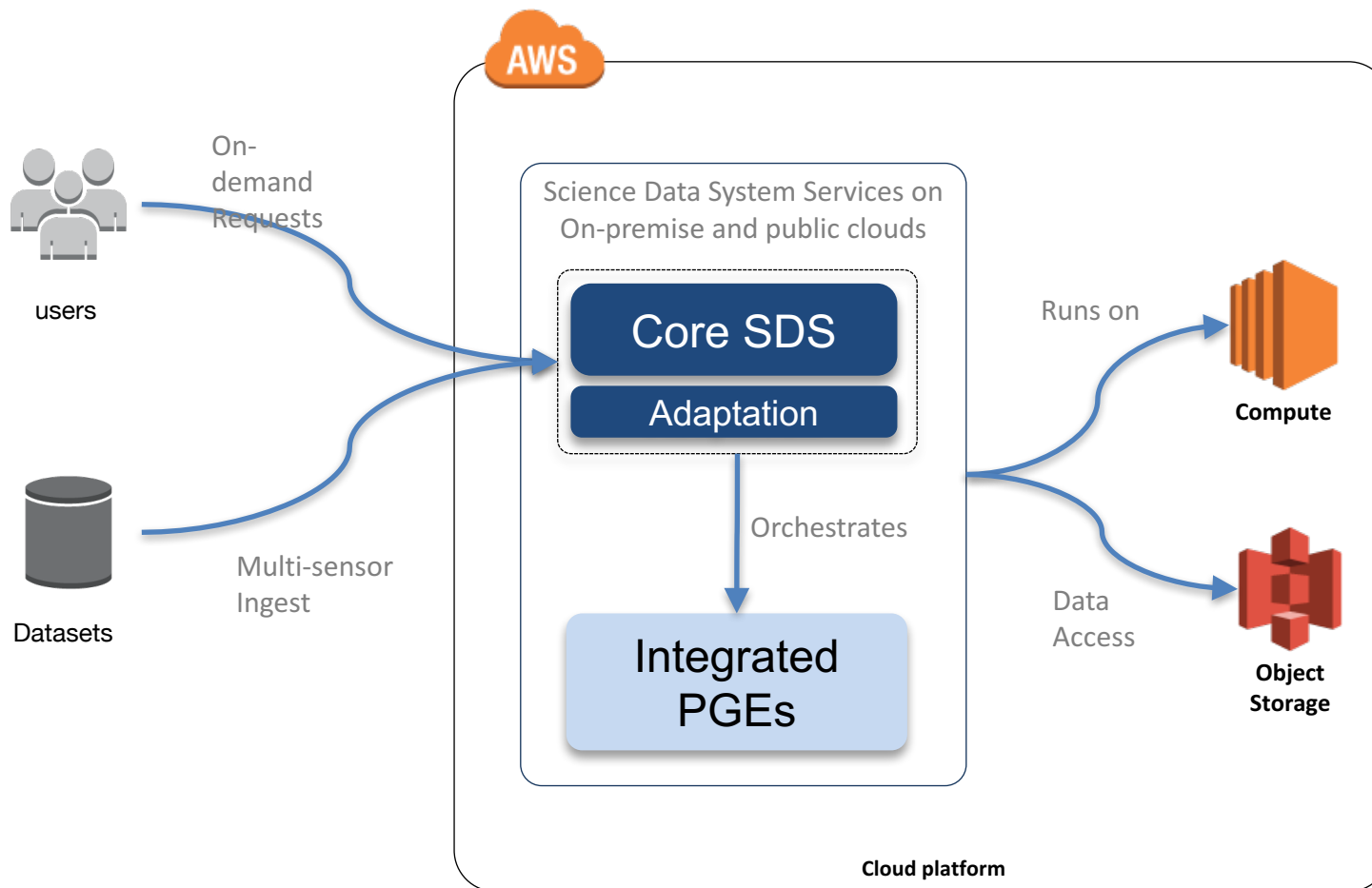
- Science data product into AWS S3 **object storage**
- **Scale up** compute nodes to run in AWS EC2
- **Internal** SDS data throughput needs are scalable via **cloud architecture**
  - Object storage can scale up **data volume** and **aggregate data throughput** by compute instances
- Architectural components can be **collocated**



# Relevant Software Components for Onboarding



- SDS with domain adaptation
- Product Generation Executive (PGE) orchestration
- Data management



# Key Onboarding Steps



- Systems Engineering
- Cloud economics
  - TCO analysis
  - Deployment topology
  - Cost bracketing strategies
- PGEs Docker and workflow orchestration
  - Dockerization
  - Continuous Integration (CI)
- Data management (ingest, metadata handling)
- Validation of cloud variant
- Benchmarking
- Large-scaling validation
- Iterate..

- Characterizing the processing domain
  - Data products per day, jobs per job, data rate needs
- Mapping to cloud model
  - Deployment topology
  - Compute instance types
  - Storage strategies
  - Network egress implications
- IT Security
  - Compliance
  - Reach back to institution services needed? (e.g. LDAP)



# Characterizing Processing Needs



- Benchmark PGE characteristics

ec2 instance	soundings per job	Average job wall-clock time (seconds)
c3.4xlarge	32	1,446
c3.4xlarge	64	2414.430
c3.4xlarge	112	4034.528
c3.4xlarge	160	5758.301
c3.8xlarge	32	888.693
c3.8xlarge	128	2473.234
c3.8xlarge	224	4073.497
c3.8xlarge	320	5473.847
r4.4xlarge	32	1473.266
r4.4xlarge	64	3878.281
r4.4xlarge	112	3878.281
r4.4xlarge	160	5547.405
r4.8xlarge	32	971.131
r4.8xlarge	128	2786.829
r4.8xlarge	224	4527.259
r4.8xlarge	320	5836.706

# Characterizing Processing Needs



- Estimating Processing Schedule

week		1							
day		1	2	3	4	5	6	7	8
jobs									
	c3.4xlarge	10923	10923	10923	10923	10923	10923	10923	10923
	c3.8xlarge	8713	8713	8713	8713	8713	8713	8713	8713
	r4.4xlarge	2399	2399	2399	2399	2399	2399	2399	2399
	r4.8xlarge	1125	1125	1125	1125	1125	1125	1125	1125
	total	23160	23160	23160	23160	23160	23160	23160	23160

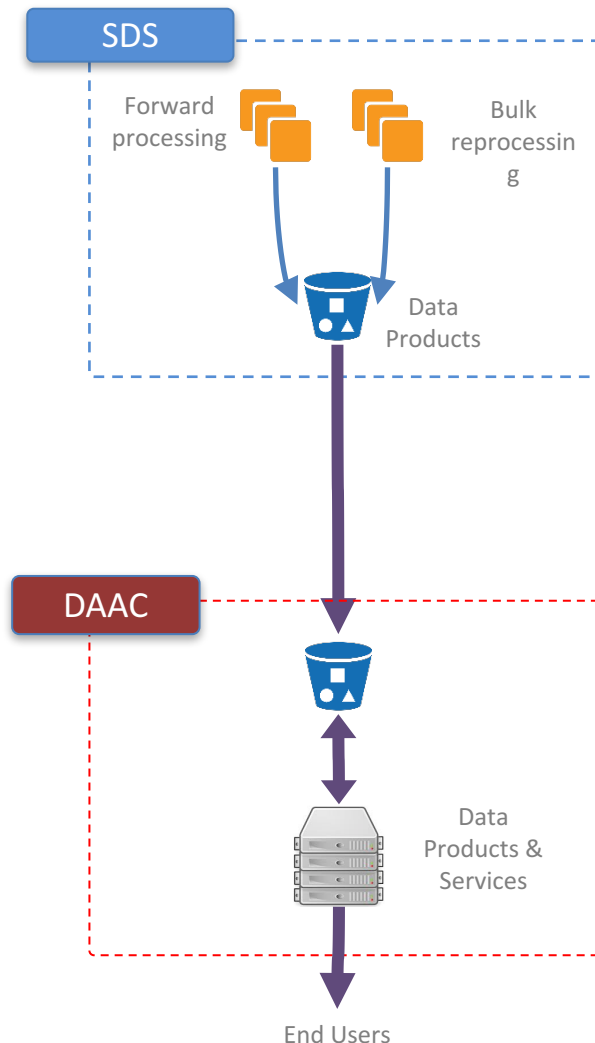
# Mapping to Cloud Resources



- Jobs to Compute to Science Measurements

week		1						
day		1	2	3	4	5	6	7
Nodes								
	c3.4xlarge	364	364	364	364	364	364	364
	c3.8xlarge	276	276	276	276	276	276	276
	r4.4xlarge	77	77	77	77	77	77	77
	r4.8xlarge	38	38	38	38	38	38	38
	total	755	755	755	755	755	755	755
hours								
	c3.4xlarge	8736	8736	8736	8736	8736	8736	8736
	c3.8xlarge	6624	6624	6624	6624	6624	6624	6624
	r4.4xlarge	1848	1848	1848	1848	1848	1848	1848
	r4.8xlarge	912	912	912	912	912	912	912
	total	18120	18120	18120	18120	18120	18120	18120

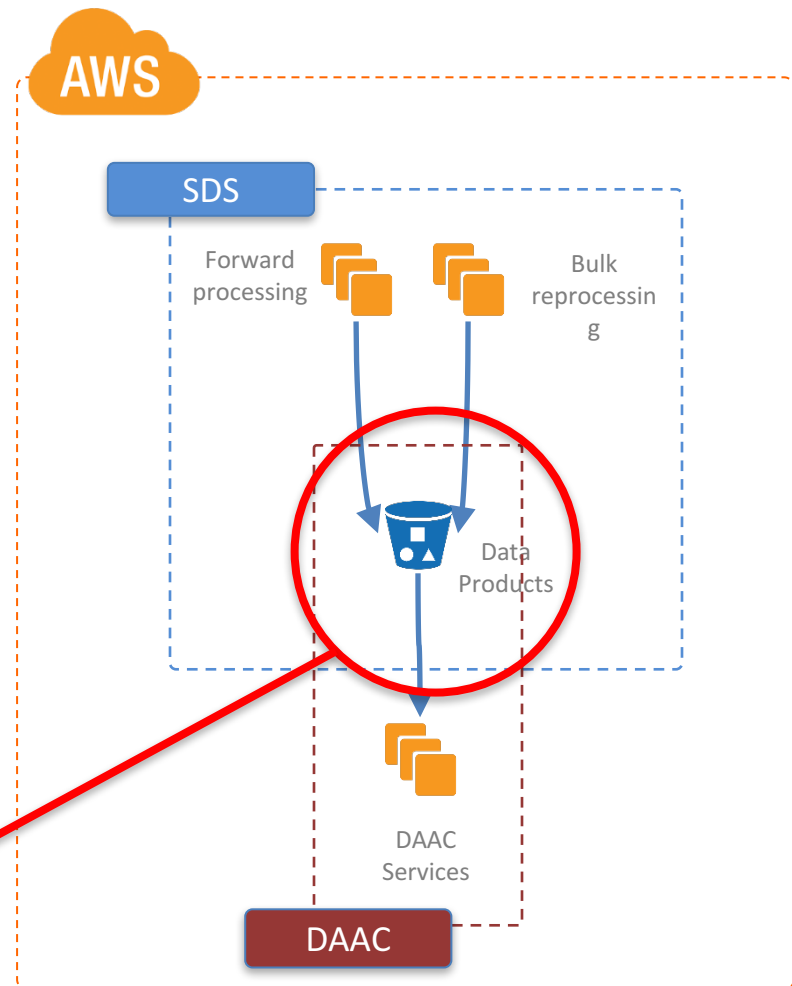
# Classic Deployment of SDS and DAAC



- Science data product generation at SDS
- Science data products moved to DAAC facilities
  - (copying large data volumes)
- End users access from DAAC
- Bottlenecks and cost impact of high network data stream

- **Shared data as interface** between SDS and DAAC
- No **egress** nor external network limitations between SDS and DAAC
- DAAC still incurs end-user egress costs.

Shared data storage



# Total Cost of Ownership (TCO)



- Factoring in compute, storage, network, topology, storage tiering, etc.
- Example monthly rollup for forward stream processing TCO in AWS
  - Does not show other costs e.g. cloud development
  - (This example uses public “rack rates”)*

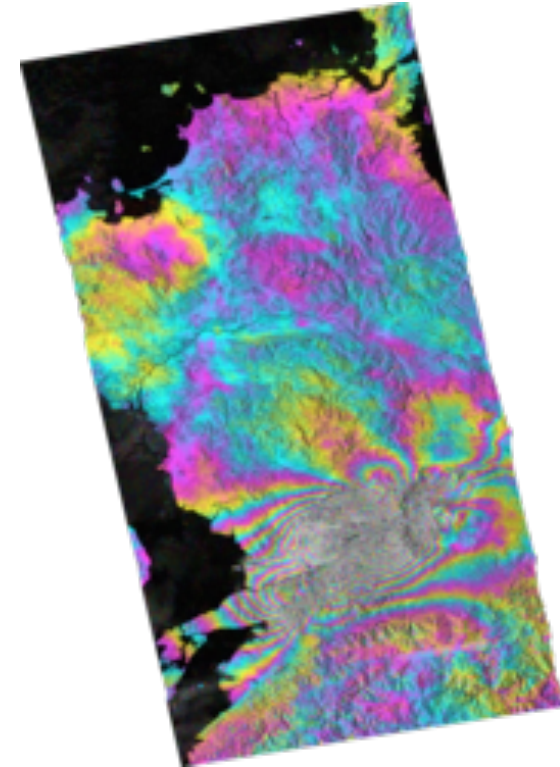
event	processing month	Compute					forward stream				Data Transfer		
							Mission Storage		Hot Data Storage				
		m4.4xlarge (#)	m4.4xlarge (\$)	r3.8xlarge (#)	r3.8xlarge (costs)	Total (\$)	S3 IA (TB)	Costs (\$)	S3 (TB)	Costs (\$)	To DAAC (TB)	Data Out Costs (\$) – if out of AWS	DX Charge
launch	-1	54	\$4,346.78	98	\$36,634.75	\$40,981.54	300	\$3,840.00	100	\$3,100.00	0	\$0.00	\$1,620.00
	0	54	\$4,346.78	98	\$36,634.75	\$40,981.54	300	\$3,840.00	451	\$13,973.29	0	\$0.00	\$1,620.00
	1	54	\$4,346.78	98	\$36,634.75	\$40,981.54	398	\$5,088.00	802	\$24,846.50	351	\$32,293.55	\$1,620.00
	2	54	\$4,346.78	98	\$36,634.75	\$40,981.54	495	\$6,336.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	3	54	\$4,346.78	98	\$36,634.75	\$40,981.54	593	\$7,584.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	4	54	\$4,346.78	98	\$36,634.75	\$40,981.54	690	\$8,832.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	5	54	\$4,346.78	98	\$36,634.75	\$40,981.54	788	\$10,080.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	6	54	\$4,346.78	98	\$36,634.75	\$40,981.54	885	\$11,328.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	7	54	\$4,346.78	98	\$36,634.75	\$40,981.54	983	\$12,576.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	8	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1080	\$13,824.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	9	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1178	\$15,072.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	10	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1275	\$16,320.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	11	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1373	\$17,568.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	12	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1470	\$18,816.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	13	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	14	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	15	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	16	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	17	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	18	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	19	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	20	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	21	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	22	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	23	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00
	24	54	\$4,346.78	98	\$36,634.75	\$40,981.54	1568	\$20,064.00	1112	\$34,472.00	351	\$32,293.55	\$1,620.00



# Example Cost Model for Sentinel-1A/B Production



- Processing to 1000 x Sentinel-1A Level-2 phase unwrapped interferograms
  - Cost examples based on published AWS rack rates



Compute					
	per scene processing (m)	total (hr)	ec2 costs (\$)	EBS storage (GB)	EBS costs (\$)
L1 IW_SLC	5	83.3	\$14.00	500	\$6.94
L1 IW_SLC_SWATH	5	500.0	\$84.00	500	\$41.67
L2 interferogram	300	12500.0	\$2,100.00	500	\$1,041.67
	310.0	13083.3	\$2,198.00	1500.00	\$1,090.28

Storage				Access			
	scenes	volume (GB)	s3-ia monthly costs (\$)	data retrieved (%)	data egress (TB)	s3-ia data retrieval monthly costs (\$)	egress monthly costs (\$)
L1 IW_SLC	1000	283.0	\$5.66	10%	0.03	\$0.28	\$4.39
L1 IW_SLC_SWATH	6000	1698.0	\$33.96	10%	0.17	\$1.70	\$26.32
L2 interferogram	2500	6250.0	\$125.00	100%	6.10	\$62.50	\$968.75
	9500	8231.0	\$164.62		184.63	\$64.48	\$999.46

# Benchmarking for Optimizing Costs



- Leveraging AWS spot market
- Optimize bidding strategies:
  - Bid high threshold to avoid compute node terminations
  - *Big low to run cheap (but with disruptions)*

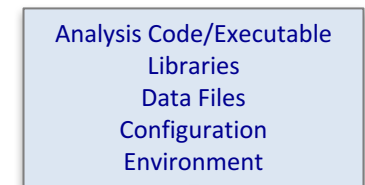
	Input
ec2 instance	soundings per job
c3.4xlarge	32
c3.4xlarge	64
c3.4xlarge	112
c3.4xlarge	160
c3.8xlarge	32
c3.8xlarge	128
c3.8xlarge	224
c3.8xlarge	320
r4.4xlarge	32
r4.4xlarge	64
r4.4xlarge	112
r4.4xlarge	160
r4.8xlarge	32
r4.8xlarge	128
r4.8xlarge	224
r4.8xlarge	320

# “Containerizing” PGEs



- **Containerizing**
  - Encapsulating analysis steps into more portable and self-contained Docker Containers
- **Agility**
  - Foster agility through rapid development and deployment of analysis steps
- **Portability**
  - Deploy analysis steps in private and public clouds
- **Scalability**
  - Large-scale deployment of Containers to compute fleet
- **Provenance**
  - Archive PGE Containers in AWS/S3
  - Reproduce all existing and prior versions of data analysis and production
  - “use what you store, and store what you use”
  - Re-run analysis by data system and DAAC

*“Docker containers wrap up a piece of software in a complete filesystem that contains everything it needs to run: **code, runtime, system tools, system libraries** – anything you can install on a server. This guarantees that it will always run the same, regardless of the environment it is running in.”*



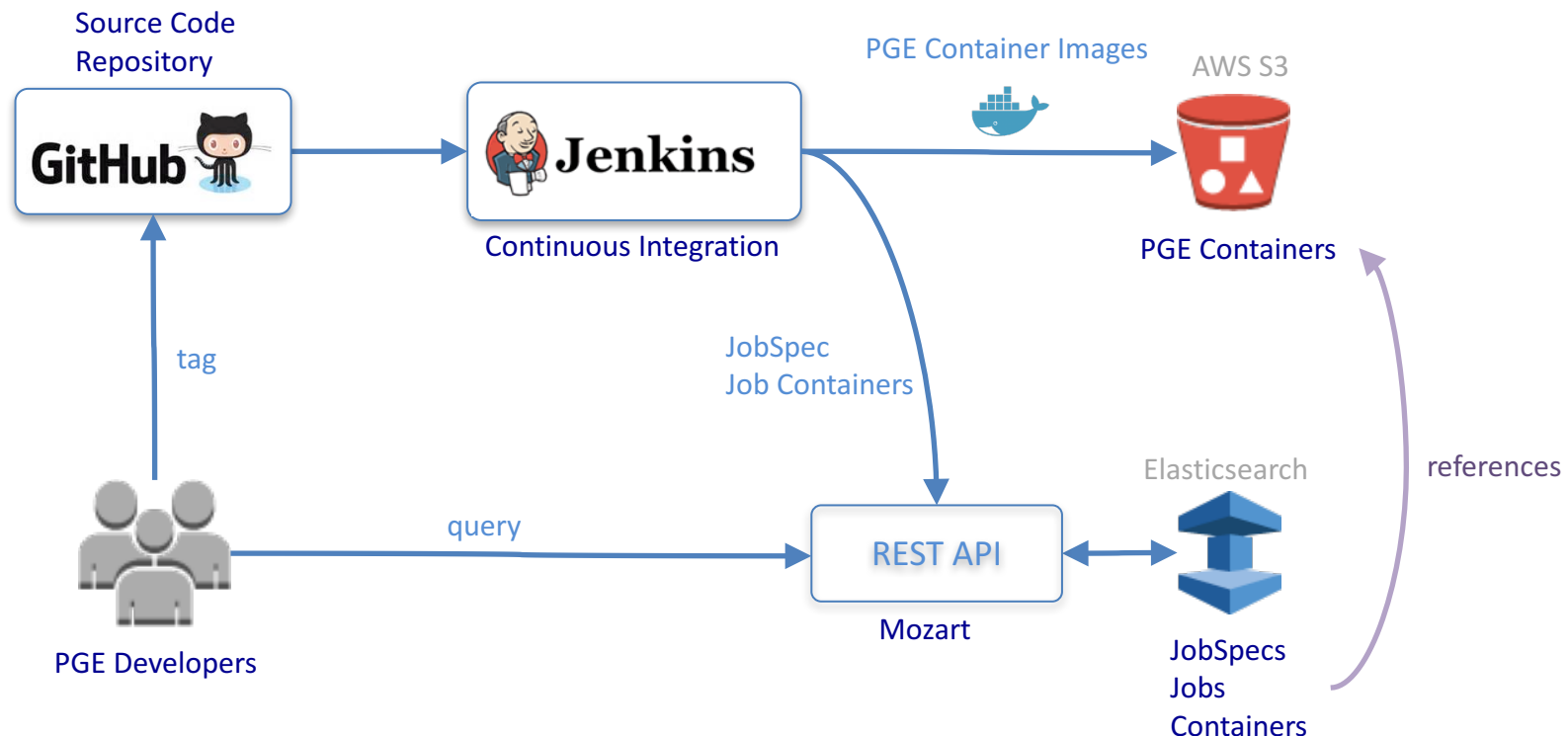
Export to / load from  
container tarballs in  
AWS/S3



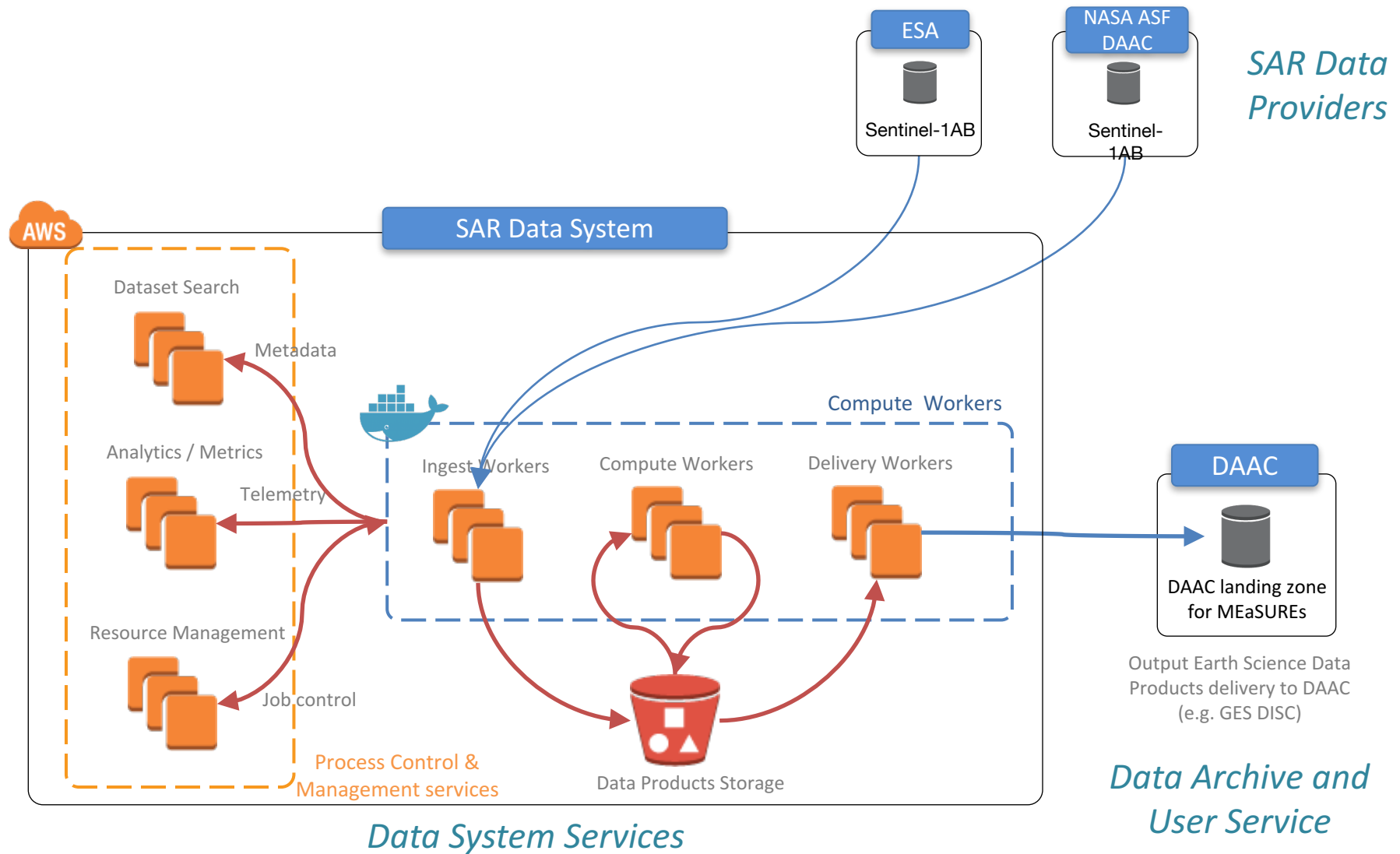
# Continuous Integration (CI) of PGE Deployment



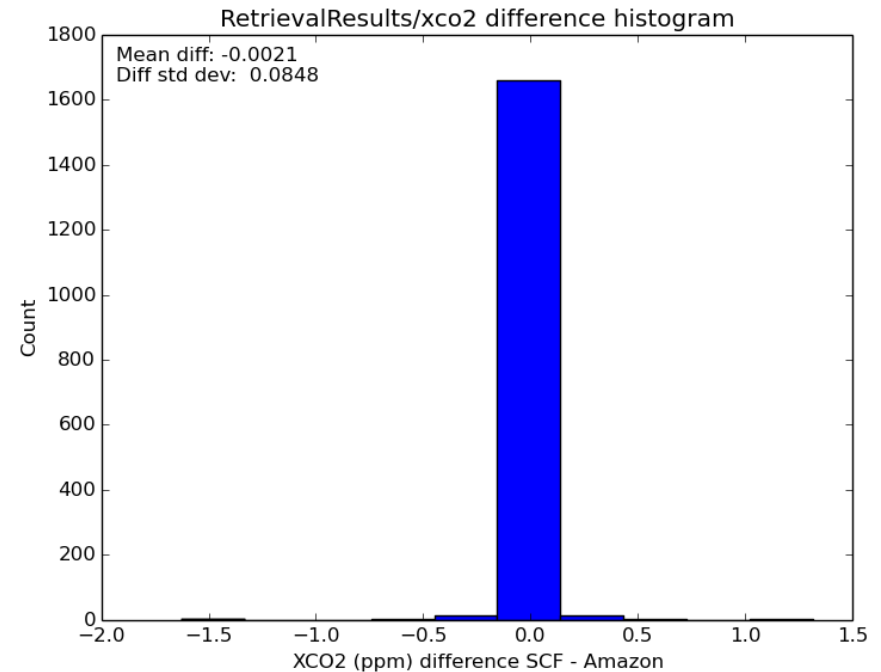
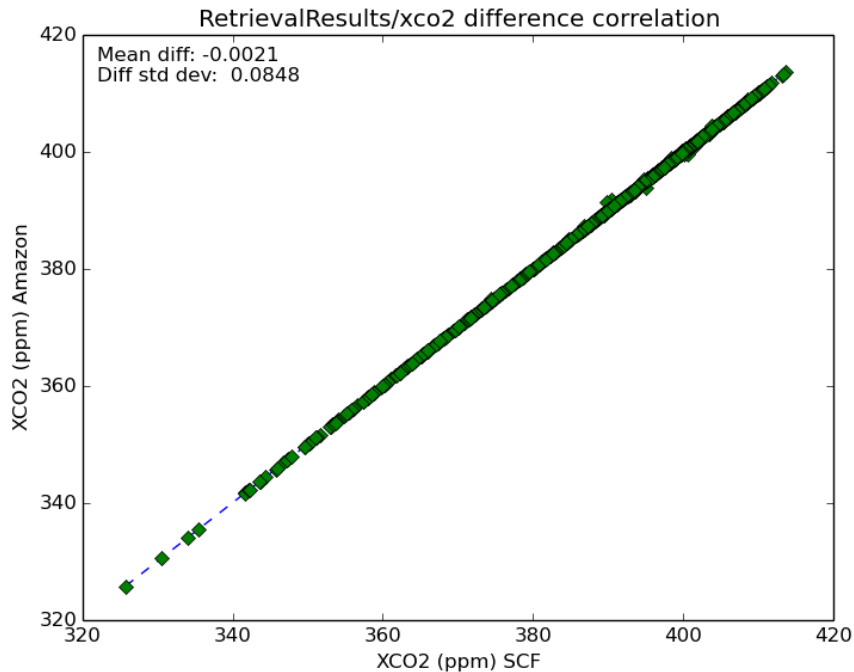
- All processing steps are jobs
  - PGEs are examples of jobs
- All job environments are encapsulated in Docker Containers
- JobSpec defines the job specification
  - References which Job Container holds the job environment
- A job container may contain more than one job



# End-to-End Integration



- Science and algorithm teams to validation the cloud-native version within acceptable tolerance





# Large-Scale Considerations



- Compute terminations
  - Spot market terminations
  - Availability Zone (AZ) load rebalancing
  - Instance failures
- “Job drain”
  - Addressing failures leading to job drain from work queues
- “Thundering herd”
  - API rate limit exceeded
- AWS Spot “Market Maker”
  - You affecting spot market prices
- S3 object store performance optimizations needed
- Auto-scaling
  - slow scale-up needs AWS tweaks
  - scale-down group vs self-terminating instances

# Key Points



- Cloud use can be cheaper if can dive deeper in architecture design and TCO impacts
- Cost implications of Earth Science Data Systems in the Cloud:
  - Compute, Storage, Network, Deployment Topology
- **Cloud systems engineering**
- **Cloud economics**
  - Total Cost of Ownership (TCO) analysis
- At large scales, need to deal with **scaling issues**.
- Running on **spot market** for cost savings—but need **resiliency**
- **Fault tolerant** science data systems can scale better in cloud computing environments
- **Collocation (e.g. Data Lakes)**
- **Benchmark at full-scale!**
  - Test in full-scale production
  - Assess steady-state at full-scale
  - Monitor real-time metrics
- Validate the “cloud version”



# BACKUP

# Running Job Containers on Workers



- Jobs are codified in Job Containers (Docker).
- Workers pop a Job off from Resource Manager
- Workers pull (as needed) PGE Container
- Workers run the
  - In the job info, it describes which Docker Container to localize from a repository (could be S3). Also in the job info are the job description information of how to run a job in a Container.
  - The worker then pulls down a Job Container image and execute jobs.
  - A Container can host more than one job executable.
  - Workers send back telemetry information to Mozart service
- Compute instances run the Workers.
  - Running on each compute instance is a `job_worker.py` and/or `task_worker.py`. Each of these workers employ Celery for task management and RabbitMQ for broker transport.

