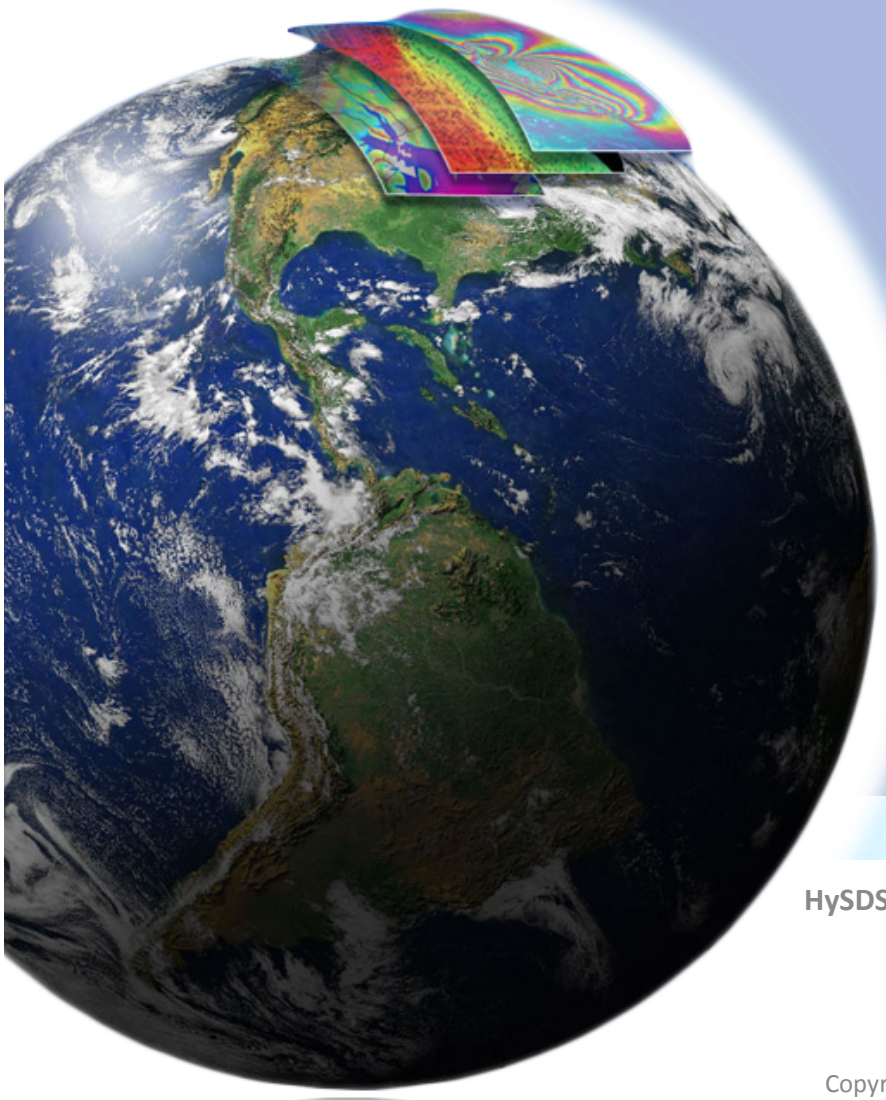


Jet Propulsion Laboratory
California Institute of Technology

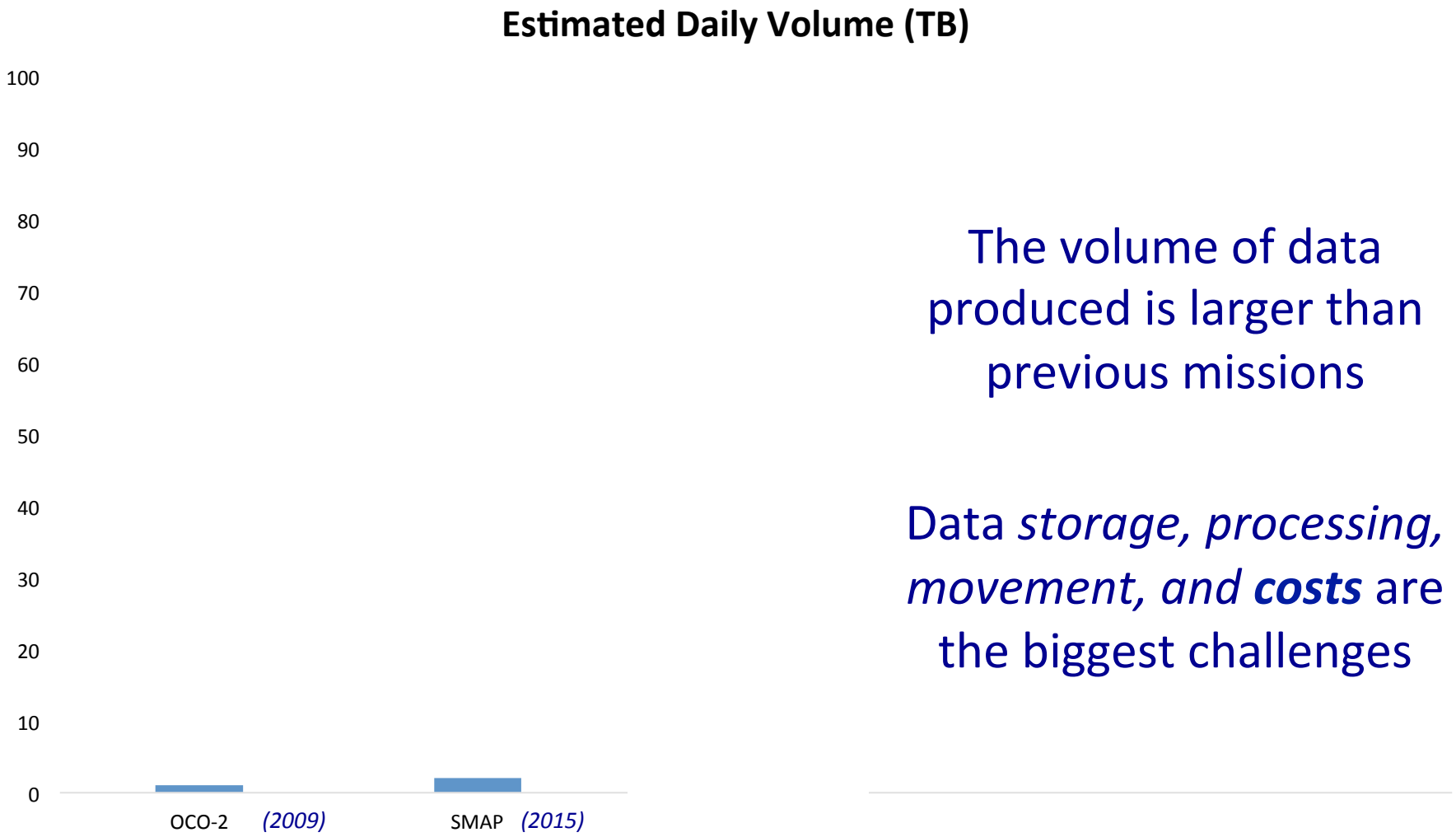
Architecting for the Cloud: Best Practices and Cost Analysis

Science Data Systems

Wednesday, January 11, 2017
ESIP Federation 2017 Winter Meeting
Bethesda, MD

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

Next Generation Missions



The volume of data produced is larger than previous missions

Data *storage, processing, movement, and* ***costs*** are the biggest challenges

Cloud Economics

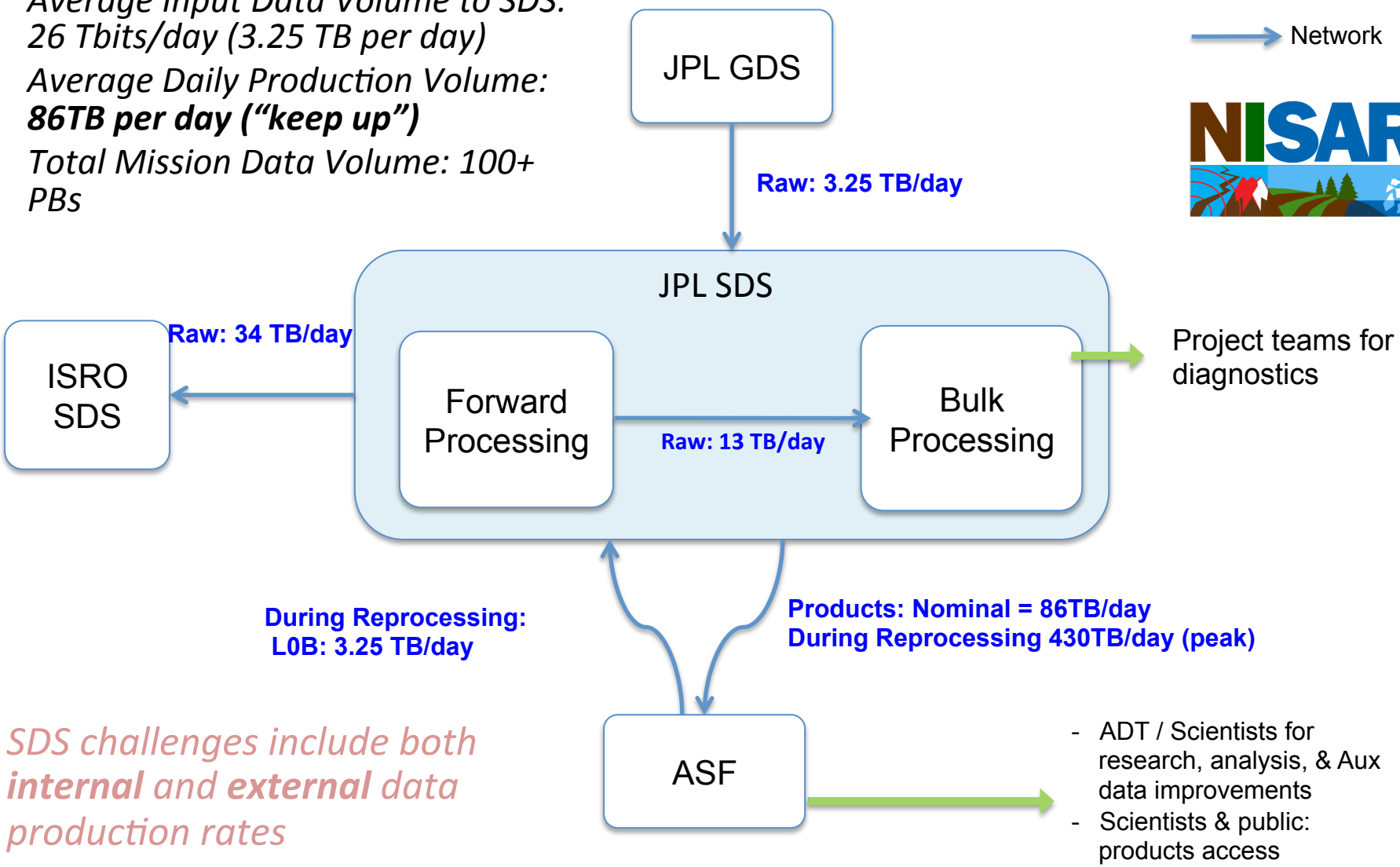
Total Cost of Ownership (TCO) implications of Earth Science Data Systems in the Cloud:

- Compute
- Storage
- Network
- Deployment Topology

- Discussion in the context of AWS.
- This is not an endorsement of any one specific cloud vendor.

NISAR Science Data System (SDS)

- *Average Input Data Volume to SDS: 26 Tbits/day (3.25 TB per day)*
- *Average Daily Production Volume:* ***86TB per day ("keep up")***
- *Total Mission Data Volume: 100+ PBs*

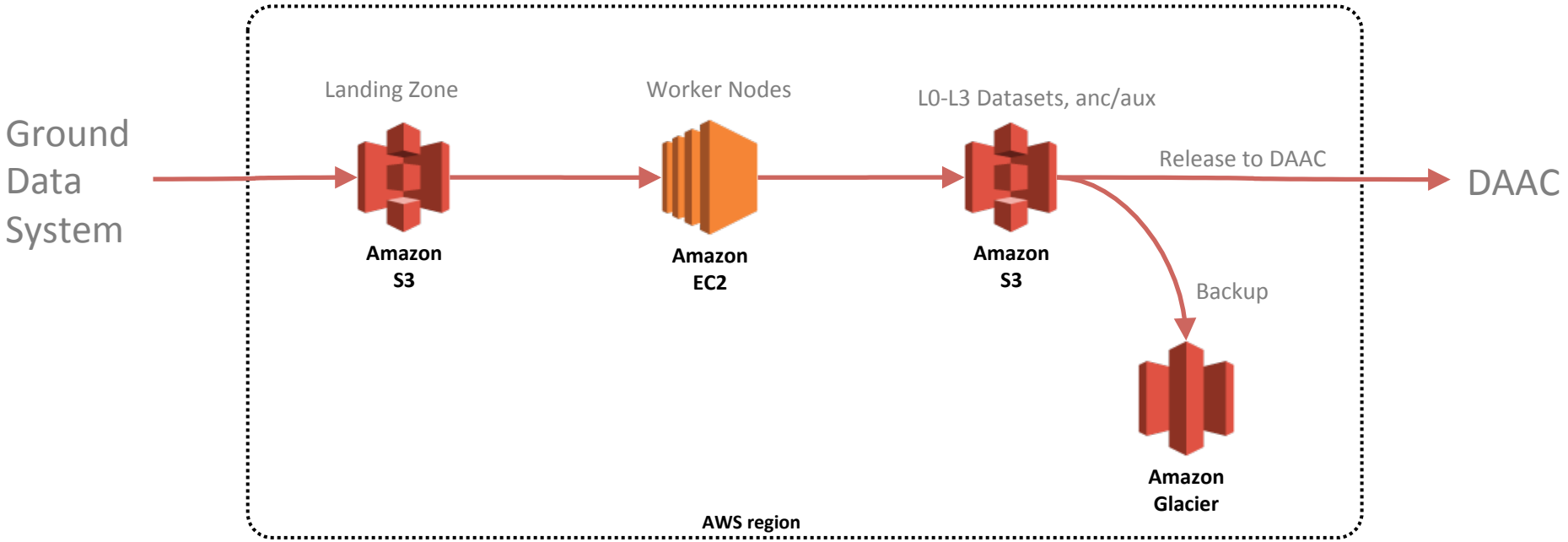


SDS challenges include both ***internal*** *and* ***external*** *data production rates*

A CLOUD-BASED APPROACH

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**



Cloud Economics

COMPUTE

AWS EC2 Instance Variations

Name	API Name	Memory	Compute Units (ECU)	vCPUs	Storage	Arch	Network Performance	EBS Optimized: Max Bandwidth	VPC Only	Linux On Demand cost	Linux Rese
C3 High-CPU Eight Extra Large	c3.8xlarge	60.0 GB	108 units	32 vCPUs	640.0 GB (2 * 320.0 GB SSD)	64-bit	10 Gigabit	N/A	No	$1.680 hourly	$1.168 hourl
C3 High-CPU Quadruple Extra Large	c3.4xlarge	30.0 GB	55 units	16 vCPUs	320.0 GB (2 * 160.0 GB SSD)	64-bit	High	2000.0 Mbps	No	$0.840 hourly	$0.584 hourl
C4 High-CPU Eight Extra Large	c4.8xlarge	60.0 GB	132 units	36 vCPUs	0 GB (EBS only)	64-bit	10 Gigabit	4000.0 Mbps	Yes	$1.591 hourly	$1.130 hourl
C4 High-CPU Quadruple Extra Large	c4.4xlarge	30.0 GB	62 units	16 vCPUs	0 GB (EBS only)	64-bit	High	2000.0 Mbps	Yes	$0.796 hourly	$0.565 hourl
Cluster Compute Eight Extra Large	cc2.8xlarge	60.5 GB	88 units	32 vCPUs	3360.0 GB (4 * 840.0 GB)	64-bit	10 Gigabit	N/A	No	$2.000 hourly	$1.090 hourl
Cluster GPU Quadruple Extra Large	cg1.4xlarge	22.5 GB	33.5 units	16 vCPUs	1680.0 GB (2 * 840.0 GB)	64-bit	10 Gigabit	N/A	No	$2.100 hourly	unavailable
D2 Double Extra Large	d2.2xlarge	61.0 GB	28 units	8 vCPUs	12000.0 GB (6 * 2000.0 GB)	64-bit	High	1000.0 Mbps	No	$1.380 hourly	$0.804 hourl
D2 Eight Extra Large	d2.8xlarge	244.0 GB	116 units	36 vCPUs	48000.0 GB (24 * 2000.0 GB)	64-bit	10 Gigabit	4000.0 Mbps	No	$5.520 hourly	$3.216 hourl
D2 Quadruple Extra Large	d2.4xlarge	122.0 GB	56 units	16 vCPUs	24000.0 GB (12 * 2000.0 GB)	64-bit	High	2000.0 Mbps	No	$2.760 hourly	$1.608 hourl
G2 Eight Extra Large	g2.8xlarge	60.0 GB	104 units	32 vCPUs	240.0 GB (2 * 120.0 GB SSD)	64-bit	10 Gigabit	N/A	No	$2.600 hourly	$1.896 hourl
General Purpose GPU 16xlarge	p2.16xlarge	732.0 GB	188 units	64 vCPUs	0 GB (EBS only)	64-bit	20 Gigabit	10000.0 Mbps	Yes	$14.400 hourly	$10.951 hour
General Purpose GPU Eight Extra Large	p2.8xlarge	488.0 GB	94 units	32 vCPUs	0 GB (EBS only)	64-bit	10 Gigabit	5000.0 Mbps	Yes	$7.200 hourly	$5.476 hourl
HI1. High I/O Quadruple Extra Large	hi1.4xlarge	60.5 GB	35 units	16 vCPUs	2048.0 GB (2 * 1024.0 GB SSD)	64-bit	10 Gigabit	N/A	No	$3.100 hourly	$1.698 hourl
High Memory Cluster Eight Extra Large	cr1.8xlarge	244.0 GB	88 units	32 vCPUs	240.0 GB (2 * 120.0 GB SSD)	64-bit	10 Gigabit	N/A	No	$3.500 hourly	$1.689 hourl
High Storage Eight Extra Large	hs1.8xlarge	117.0 GB	35 units	16 vCPUs	48000.0 GB (24 * 2000.0 GB)	64-bit	10 Gigabit	N/A	No	$4.600 hourly	$2.574 hourl
I2 Double Extra Large	i2.2xlarge	61.0 GB	27 units	8 vCPUs	1600.0 GB (2 * 800.0 GB SSD)	64-bit	High	1000.0 Mbps	No	$1.705 hourly	$0.848 hourl
I2 Eight Extra Large	i2.8xlarge	244.0 GB	104 units	32 vCPUs	6400.0 GB (8 * 800.0 GB SSD)	64-bit	10 Gigabit	N/A	No	$6.820 hourly	$3.392 hourl
I2 Quadruple Extra Large	i2.4xlarge	122.0 GB	53 units	16 vCPUs	3200.0 GB (4 * 800.0 GB SSD)	64-bit	High	2000.0 Mbps	No	$3.410 hourly	$1.696 hourl
M2 High Memory Quadruple Extra Large	m2.4xlarge	68.4 GB	26 units	8 vCPUs	1680.0 GB (2 * 840.0 GB)	64-bit	High	1000.0 Mbps	No	$0.980 hourly	$0.444 hourl
M3 General Purpose Double Extra Large	m3.2xlarge	30.0 GB	26 units	8 vCPUs	160.0 GB (2 * 80.0 GB SSD)	64-bit	High	1000.0 Mbps	No	$0.532 hourly	$0.380 hourl
M4 16xlarge	m4.16xlarge	256.0 GB	188 units	64 vCPUs	0 GB (EBS only)	64-bit	20 Gigabit	10000.0 Mbps	Yes	$3.447 hourly	$2.369 hourl
M4 Deca Extra Large	m4.10xlarge	160.0 GB	124.5 units	40 vCPUs	0 GB (EBS only)	64-bit	10 Gigabit	4000.0 Mbps	Yes	$2.155 hourly	$1.480 hourl

EC2Instances.info

- Hyperthreading and vCPUs
  - m2.4xlarge instances *previously had hyperthreading disabled*
- *Multi-tenancy and contention*
  - m2.4xlarge instance previously occupied the entire physical node, which had no **multi-tenancy contention.**
- Memory to CPU ratio
- **AWS trend to move away from ephemeral/instance storage to SSD to EBS**
- Network bandwidth
- EBS bandwidth

Benchmark Use Case: OCO-2 L2 Full Physics

Benchmarks for OCO-2 Level 2 Full Physics processors

instance	vCPU	physical hardware	cores	threads	memory	memory-cpu ratio	disks	test runs: soundings per job	test runs: concurrent processors	test runs: # test runs	test runs: mean runtime (s)
m2.4xlarge	8	Intel Xeon E5-2665 @ 2.40 GHz 1 processor, 8 threads	8	8	68.4	8.55	2 x 840	32	8	42	13[illegible]2
cc2.8xlarge	32	Intel Xeon E5-2670 @ 2.59 GHz 2 processors, 16 cores, 32 threads	16	32	60.5	1.89	4 x 840	32	32	14365	615
m3.2xlarge	8	Intel Xeon E5-2670 @ 2.60 GHz 1 processor, 8 threads	4	8	30	3.75	SSD 2 x 80				
c3.8xlarge	32	Intel Xeon E5-2680 v2 @ 2.80 GHz 2 processors, 16 cores, 32 threads	16	32	60	1.88	SSD 2 x 320	32	32	126	461
r3.8xlarge	32	Intel Xeon E5-2670 v2 @ 2.49 GHz 2 processors, 16 cores, 32 threads	16	32	244	7.63	SSD 2 x 320	32	32	131	489

- m2.4xlarge has best performance for OCO-2 l2_fp processing
  - m2.4xlarge instance *[may]* occupy the entire physical node. No **multi-tenancy** contention
  - 32 concurrent processors on c3 and r3 instances may have more **memory and disk contention**
  - More memory on r3.8xlarge not as large affect on performance
  - m2.4xlarge does not have SSD disks. L2_fp may not be as heavy on disk I/O bound.
- Optimize job sizes per instance type
  - m2.4xlarge: use 80 soundings per 1-hour job
  - c3.8xlarge: use 256 soundings per 1-hour job

"Cost Per Unit" Estimates

prefered minimun job duration (m) :	50

	test runs					prefered job size		on-demand (US-West-1)		spot (US-West-1)		on-demand (US-West-2)		spot (US-West-2)	
instance	soundings per job	concurrent processors	# test runs	mean runtime (s)	mean runtime per sounding per processor (m)	# soundings	# soundings (evenly distributed processors)	mean costs per job	$ per sounding	mean costs per job	$ per sounding	mean costs per job	$ per sounding	mean costs per job	$ per sounding
m2.4xlarge	32	8	42	1382	5.8	69	7	$0.4223	0.01320	$0.0250	0.00078	$0.4138	0.01293	$0.0384	0.00120
cc2.8xlarge	32	32	14365	615	10.3	156	16					$0.3417	0.01068	$0.0462	0.00144
m3.2xlarge															
c3.8xlarge	32	32	126	461	7.7	208	22	$0.2448	0.00765	$0.2151	0.00672	$0.2151	0.00672	$0.3073	0.00960
r3.8xlarge	32	32	131	489	8.2	196	22	$0.4238	0.01324	$0.0543	0.00170	$0.3803	0.01189	$0.3803	0.01189

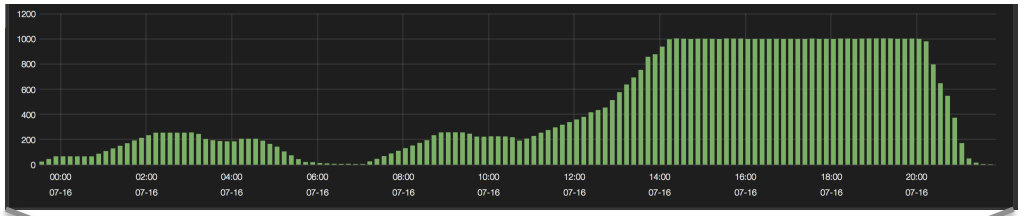
- *OCO-2's "$ per sounding" yielded non-obvious results*
  - *More* expensive compute instance was *cheaper* per science unit output

Spot Market

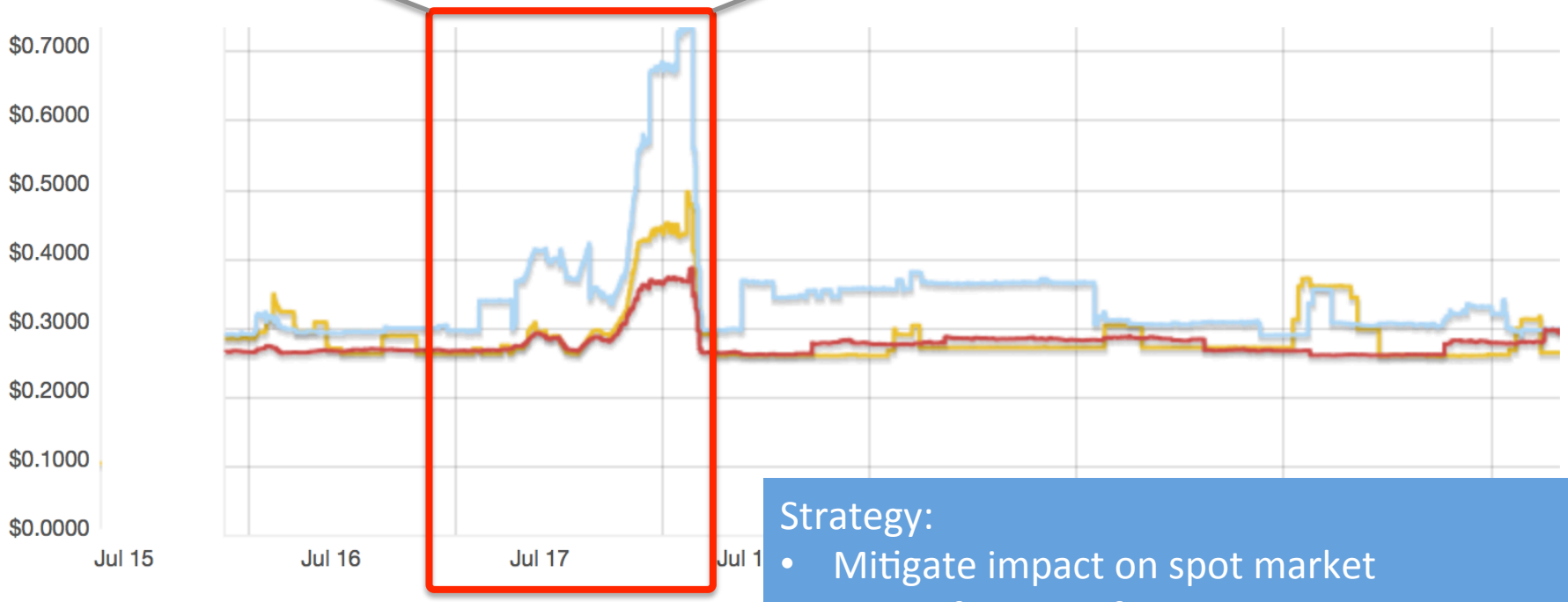
					US-West-2 (Oregon)							
					Hourly Costs				Per vCPU Costs			
instance	**vCPU**	**memory**	**memory-cpu ratio**	**disks**	**on-demand ($/hr)**	**reserved 1-yr upfront ($/hr)**	**reserved 3-yr upfront ($/hr)**	**spot linux ($/hr)**	**on-demand ($/cpu/hr)**	**reserved 1-yr upfront ($/cpu/hr)**	**reserved 3-yr upfront ($/cpu/hr)**	**spot linux ($/cpu/hr)**
m2.4xlarge	8	68.4	8.55	2 x 840	$1.0780	$0.4087	$0.2444	$0.1000	$0.1348	$0.0511	$0.0306	$0.0125
cc2.8xlarge	32	60.5	1.89	4 x 840	$2.0000	$0.9131	$0.6137	$0.2705	$0.0625	$0.0285	$0.0192	$0.0085
m3.2xlarge	8	30.0	3.75	SSD 2 x 80	$0.6160	$0.3750	$0.2300	$0.0700	$0.0770	$0.0469	$0.0288	$0.0088
c3.8xlarge	32	60.0	1.88	SSD 2 x 320	$1.6800	$0.9920	$0.6280	$2.4001	$0.0525	$0.0310	$0.0196	$0.0750
r3.8xlarge	32	244.0	7.63	SSD 2 x 320	$2.8000	$1.4860	$0.9820	$2.8000	$0.0875	$0.0464	$0.0307	$0.0875
c3.xlarge	4	7.5	1.88	SSD 2 x 40	$0.2310	$0.1370	$0.0870	$0.0353	$0.0578	$0.0343	$0.0218	$0.0088

- Major cost savings (**75%-90%** savings over on-demand)…if can use **spot instances**
- On **spot market**, AWS will **terminate** compute instances if market prices exceed bid threshold
- Need to **self-recover** from spot instance terminations.

“Market Maker”



This OCO-2 production run of 1000 x 32vCPUs affected the market prices



Strategy:

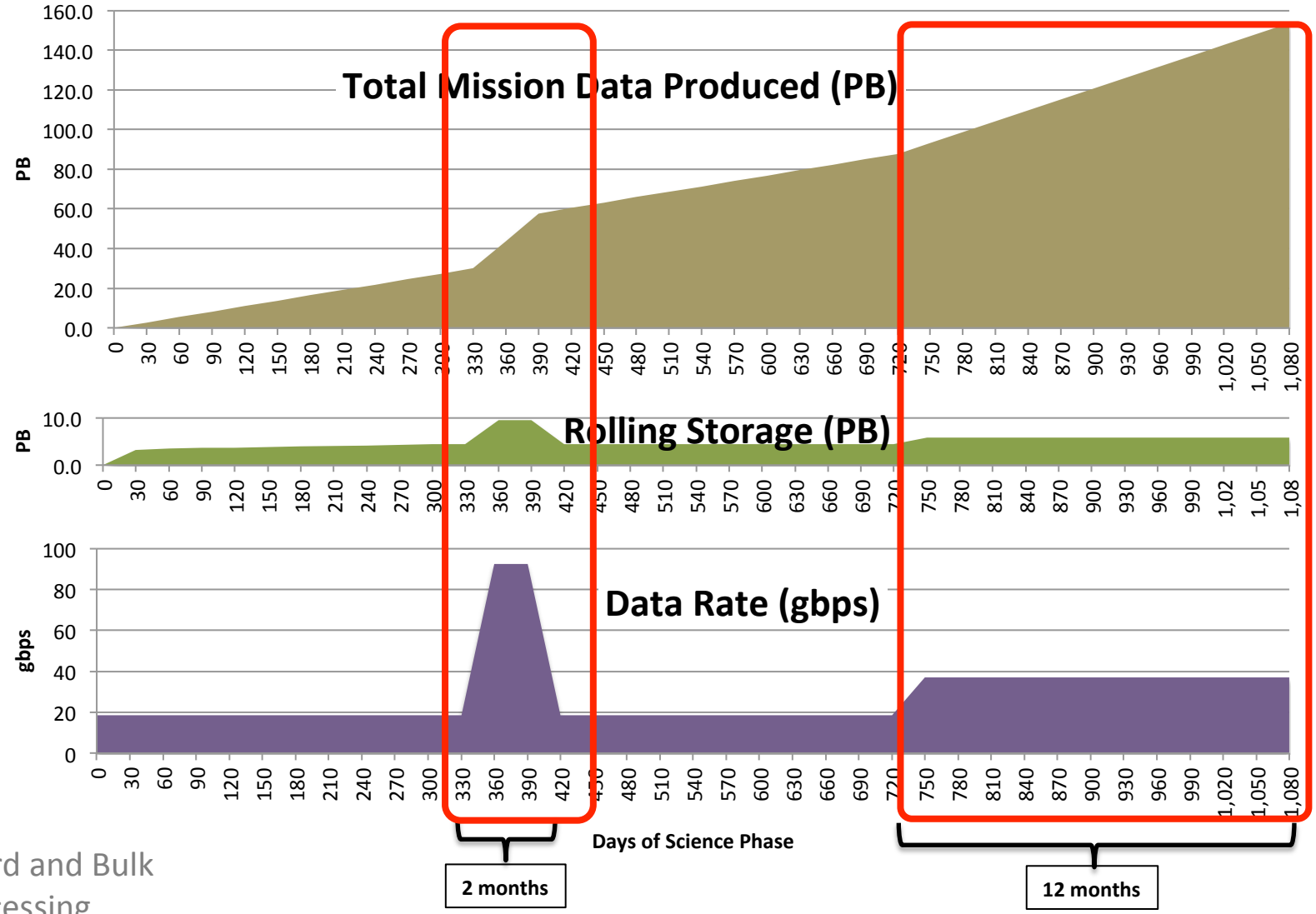
- Mitigate impact on spot market
- Diversification of resources
- “Spot fleet”

“Other” Market Makers

Cloud Economics

STORAGE

Data Volume/Throughput per Day



Forward and Bulk Reprocessing

Example Bulk Reprocessing Campaigns

AWS Storage: Monthly Tiered Example

- Storage types
  - EBS
  - S3
  - S3 Infrequent Access
  - Glacier
- Storage cost tiers
- *..Not cost-effective to store everything in S3*

Monthly S3 Storage Costs (rack rate)

At day 1

Monthly cost			Price/GB	
Tier 1: First 1 TB	1.00 TB	x	0.0300 =	30.00 $
Tier 2: Next 49 TB	49.00 TB	x	0.0295 =	1,445.50 $
Tier 3: Next 450 TB	36.00 TB	x	0.0290 =	1,044.00 **$**
Tier 4: Next 500 TB	0.00 TB	x	0.0285 =	- **$**
Tier 5: Next 4000 TB	0.00 TB	x	0.0280 =	- $
Tier 6: Above 5000 TB	0.00 TB	x	0.0275 =	- $
Total Monthly	86.00 TB			2,519.50 $

At month 1

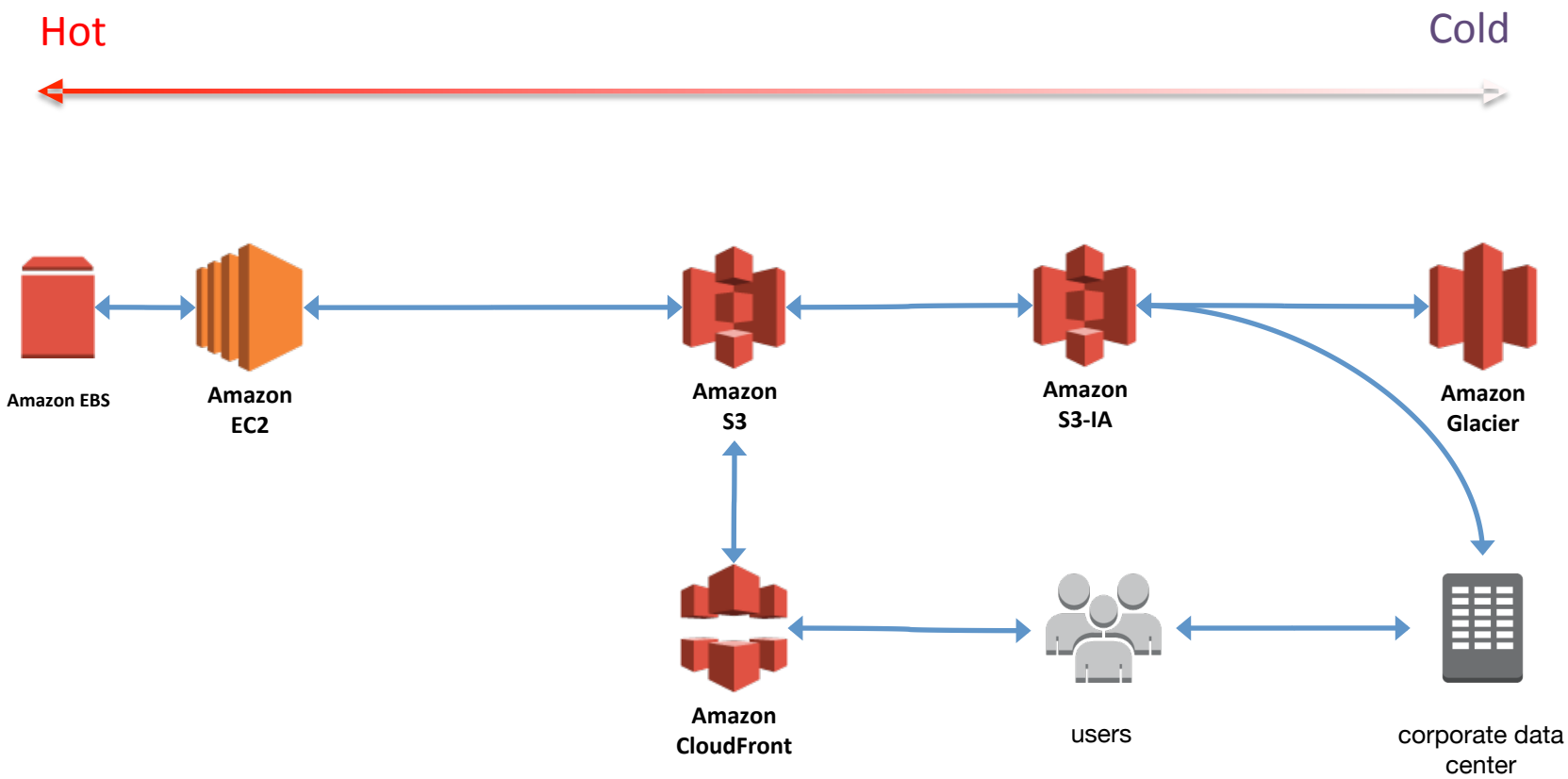
Monthly cost			Price/GB	
Tier 1: First 1 TB	1.00 TB	x	0.0300 =	30.00 $
Tier 2: Next 49 TB	49.00 TB	x	0.0295 =	1,445.50 $
Tier 3: Next 450 TB	450.00 TB	x	0.0290 =	13,050.00 **$**
Tier 4: Next 500 TB	500.00 TB	x	0.0285 =	14,250.00 **$**
Tier 5: Next 4000 TB	1580.00 TB	x	0.0280 =	44,240.00 $
Tier 6: Above 5000 TB	0.00 TB	x	0.0275 =	- $
Total Monthly	2580.00 TB			73,015.50 $

At year 1

Monthly cost			Price/GB	
Tier 1: First 1 TB	1.00 TB	x	0.0300 =	30.00 $
Tier 2: Next 49 TB	49.00 TB	x	0.0295 =	1,445.50 $
Tier 3: Next 450 TB	450.00 TB	x	0.0290 =	13,050.00 **$**
Tier 4: Next 500 TB	500.00 TB	x	0.0285 =	14,250.00 **$**
Tier 5: Next 4000 TB	4000.00 TB	x	0.0280 =	112,000.00 $
Tier 6: Above 5000 TB	26390.00 TB	x	0.0275 =	725,725.00 $
Total Monthly	31390.00 TB			866,500.50 $

Data Tiering

- Optimizing cost to storage use



Cloud Economics

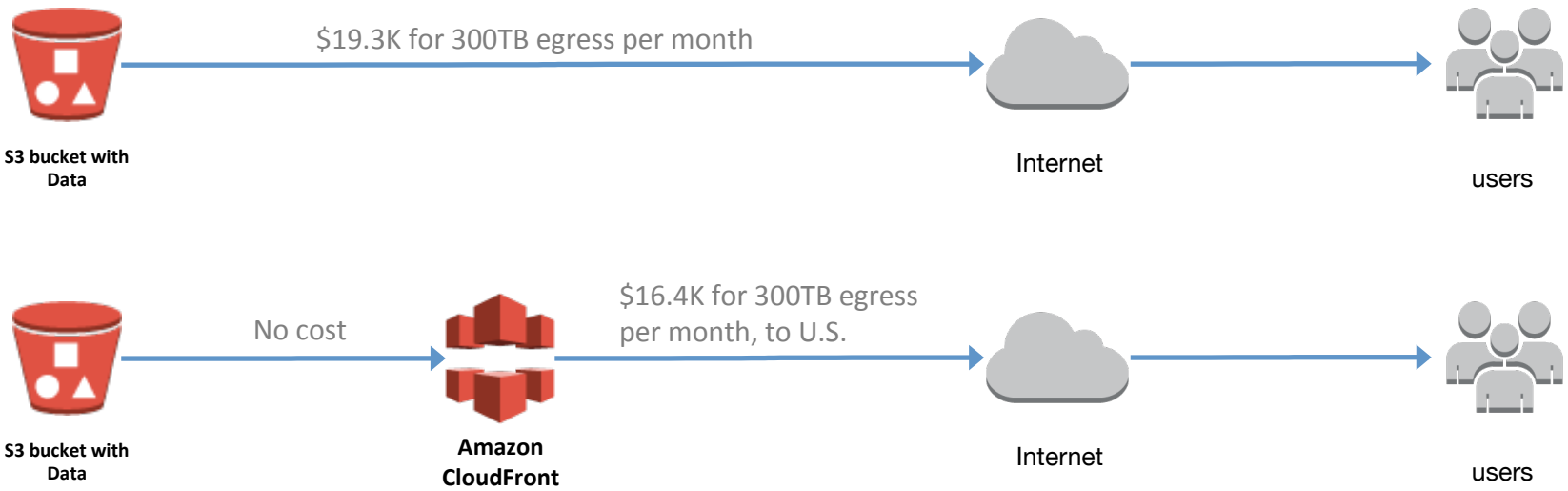
NETWORK / DATA MOVEMENT

Understanding Egress and Costs

Storage Pricing (varies by region)

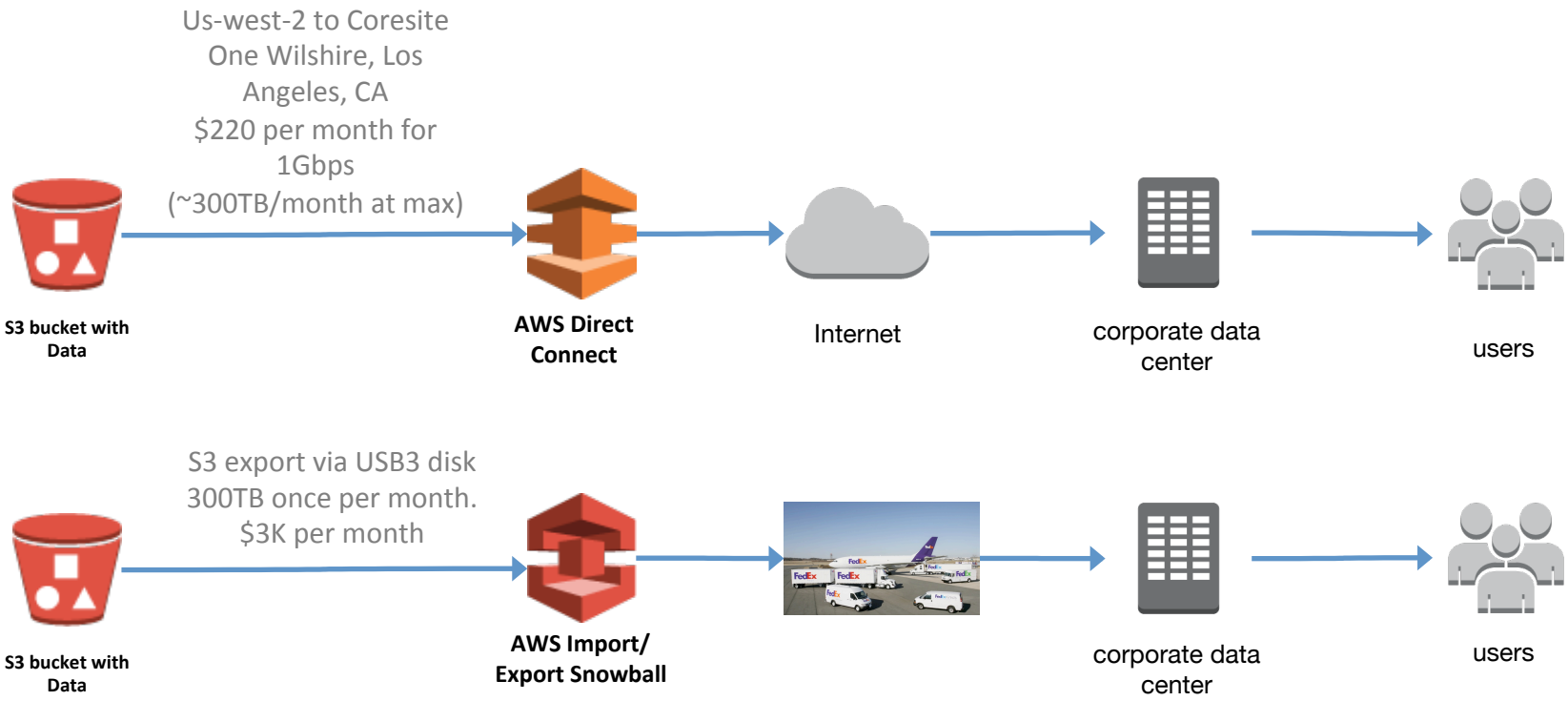
Region: US West (Oregon)

	Standard Storage	Standard - Infrequent Access Storage †	Glacier Storage
First 50 TB / month	$0.023 per GB	$0.0125 per GB	$0.004 per GB
Next 450 TB / month	$0.022 per GB	$0.0125 per GB	$0.004 per GB
Over 500 TB / month	$0.021 per GB	$0.0125 per GB	$0.004 per GB



Rate and Cost Shaping

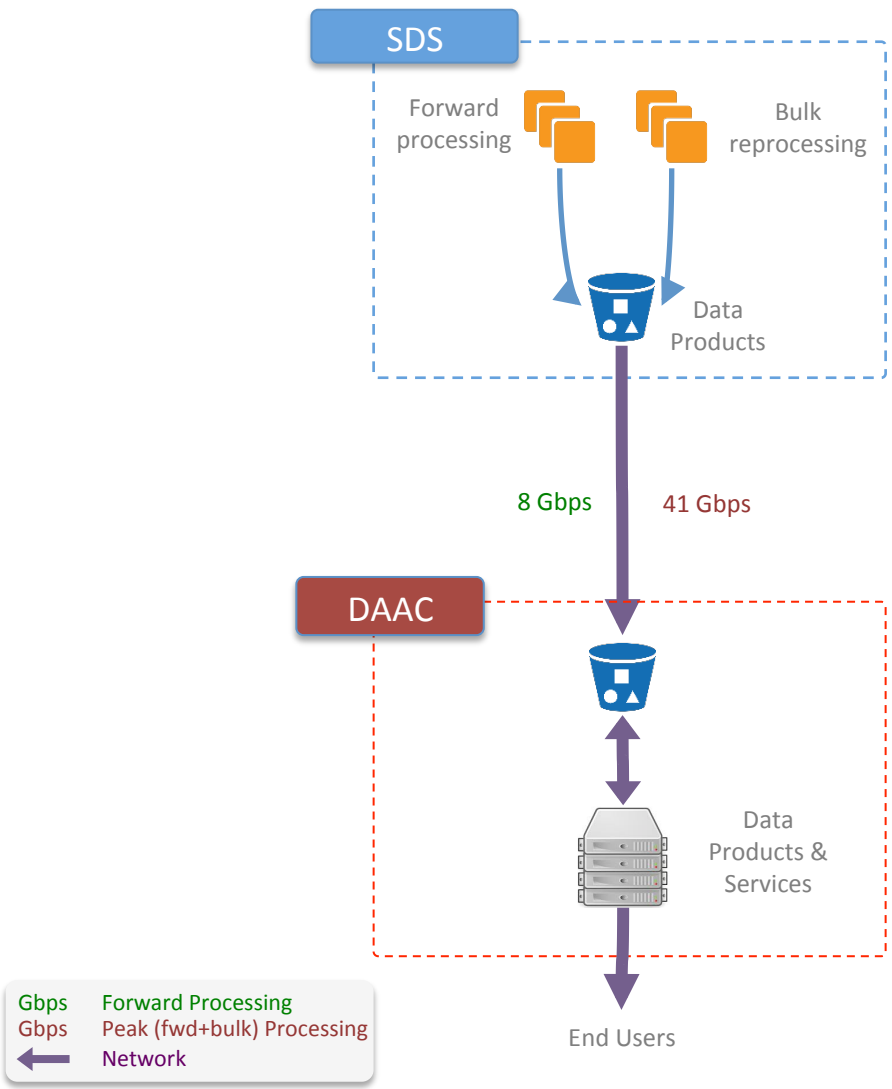
- Concern of "unbounded costs"
- Can limit bandwidth or data volume to bracket cost to bound costs
- *Note additional infrastructure and workforce costs for TCO*



Cloud Economics

DEPLOYMENT TOPOLOGY

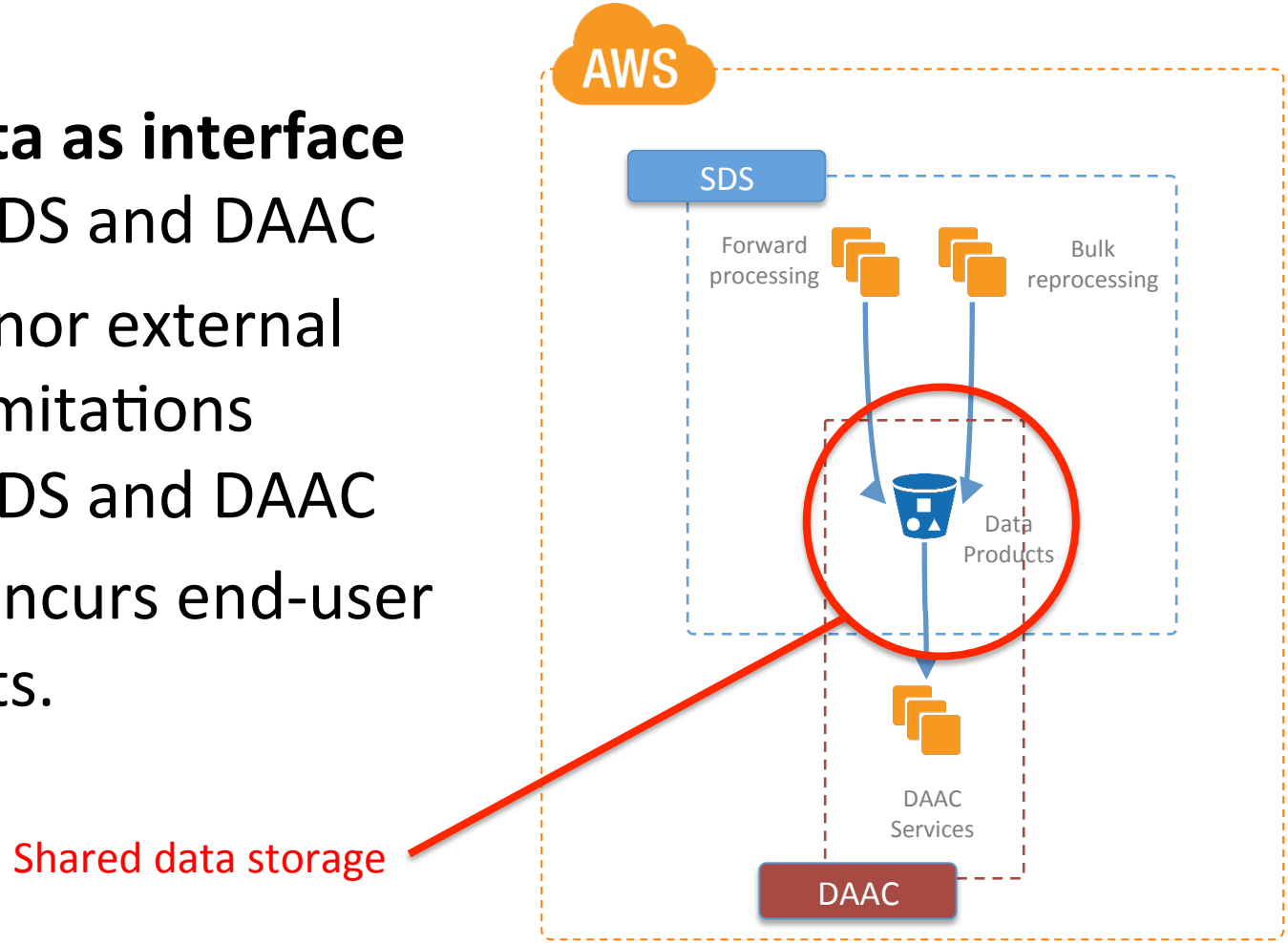
Classic Deployment of SDS and DAAC



- 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

Collocation of SDS and DAAC in Cloud

- **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.



TCO

TOTAL COST OF OWNERSHIP

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")*

		Forward Stream											
		Compute					Mission Storage		Hot Data Storage		Data Transfer		
event	processing month	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
	-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
launch	0	54	$4,346.78	98	$36,634.75	$40,981.54	300	$3,840.00	451	$13,973.25	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

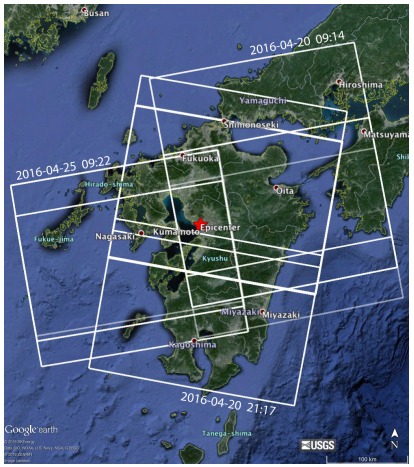
LARGE-SCALE CONSIDERATIONS

Example: SAR Analysis and Urgent Response

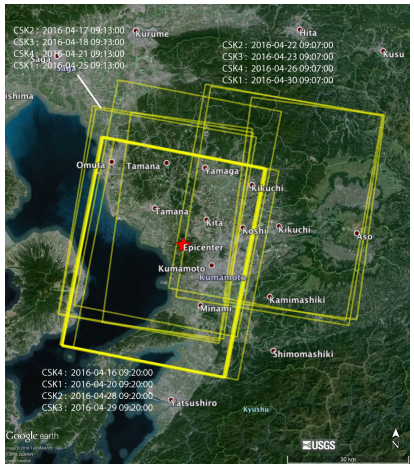
- Growing volumes of SAR Instrument data remain unprocessed to higher level science products
  - NASA Instruments: UAVSAR, AirMOSS, EcoSAR
  - ASF: ERS/ALOS/RADARSAT/ Sentinel-1 archives
  - WInSAR: Envisat archive
  - ESA: Sentinel-1 A/B
  - ASI: COSMO SkyMed archive

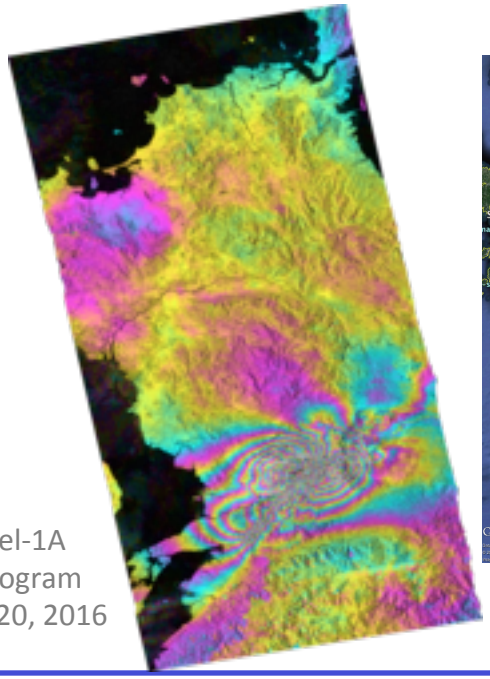
- Historically, SAR data products are hand built
  - Dependent upon specialized knowledge by investigator
  - Expensive, laborious and slow to create (days)
  - Unable to scale processing capacity

Sentinel-1A



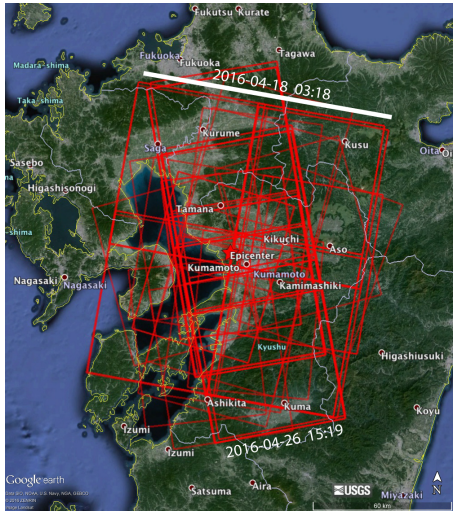
CSK





Sentinel-1A interferogram April 4 & 20, 2016

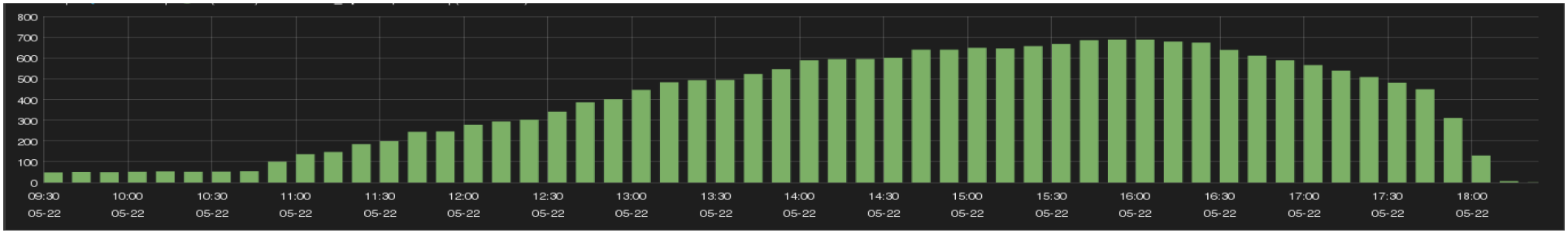
ALOS-2

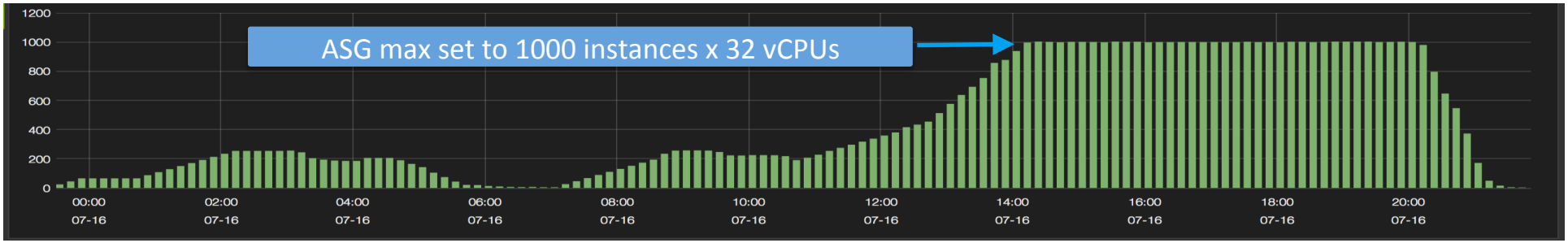


Auto-Scaling Science Data System

- *The size of the science data system compute nodes can automatically grow/ shrink based on processing demand*
- Auto Scaling group policies
  - scaling size
  - rest period
- Metric alarms
  - E.g. condition: work queue size > 20

Auto scaling enabling runs of over 100,000 ***vCPUs***





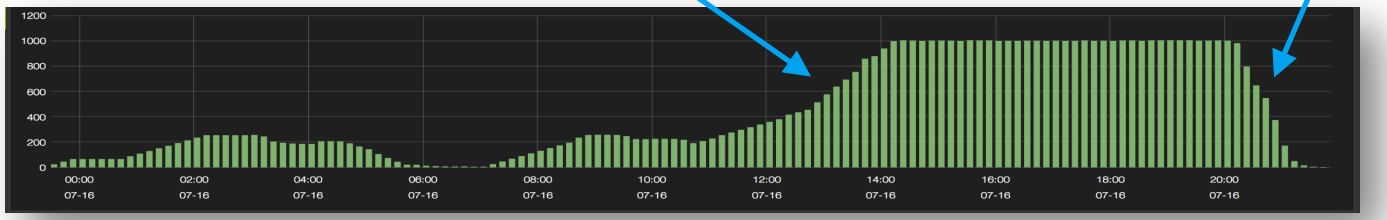
Optimizing for Scaling In/Out Events

Scaling up (scale out)

- Auto scaling group batching and timeout periods
- Scale up in group sizes of **multiples of availability zones (AZ) to m**inimize AZ load rebalancing terminations

Scaling down (scale in)

- What policy to set to scale down?
  - E.g. CPU / network utilization
- Domain knowledge only known within the compute instances
  - Self-terminating instances
  - “harakiri” / “suppuku”

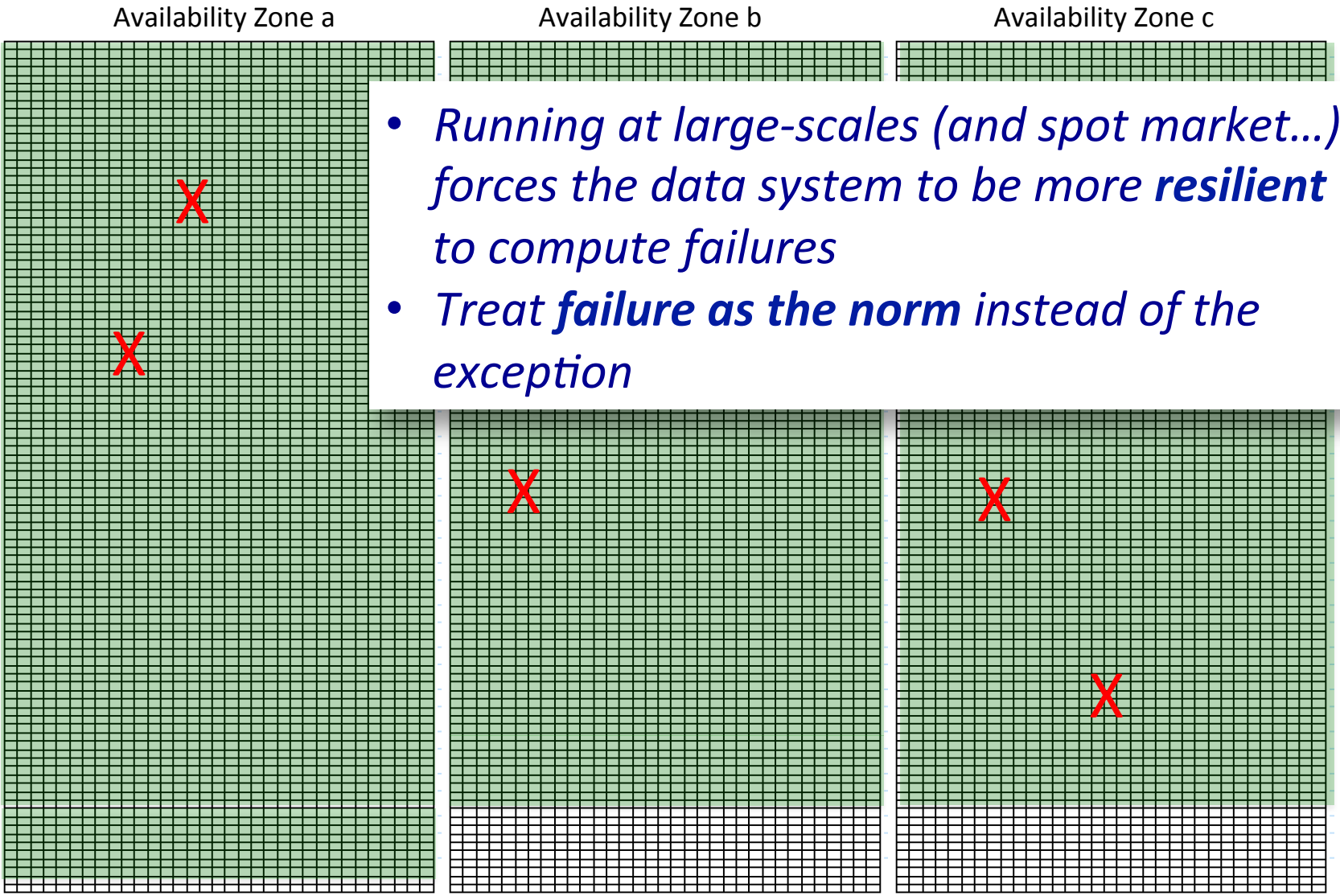


SLA Cloud Reliability

Monthly Uptime Percentage	Service Credit Percentage
Less than 99.95% but equal to or greater than 99.0%	10%
Less than 99.0%	30%

- *"The Amazon EC2 Service Level Agreement commitment is 99.95% availability for each Amazon EC2 Region."*
  - Example: expect about 1 failure per 2000 instances
    - E.g. *"EC2 has detected degradation of the underlying hardware hosting one or more of your Amazon EC2 instances in the us-west-2 region. Due to this degradation, your instance(s) could already be unreachable."*
    - E.g. disk failures, network issues, etc.
- Plan for failures !

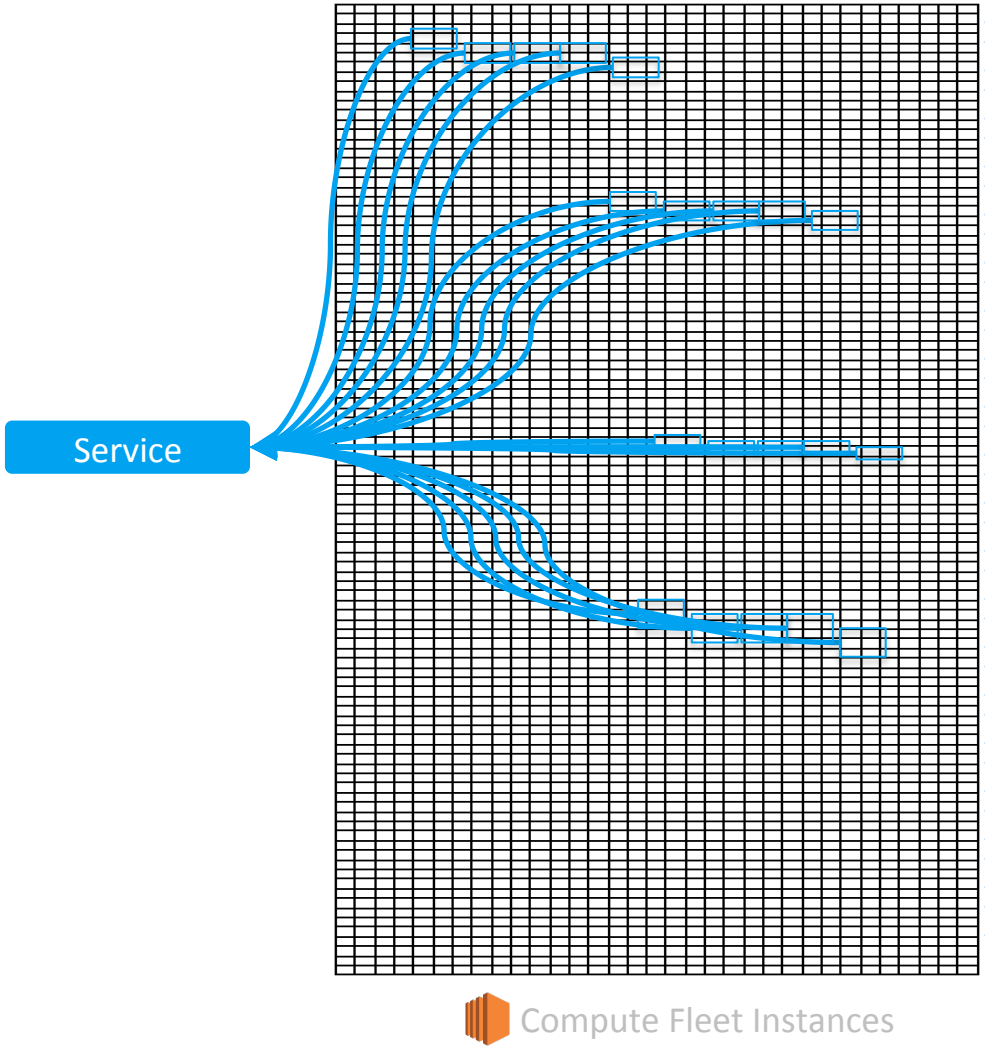
High-Resiliency at Large-Scales



- *Running at large-scales (and spot market…) forces the data system to be more **resilient** to compute failures*
- *Treat **failure as the norm** instead of the exception*

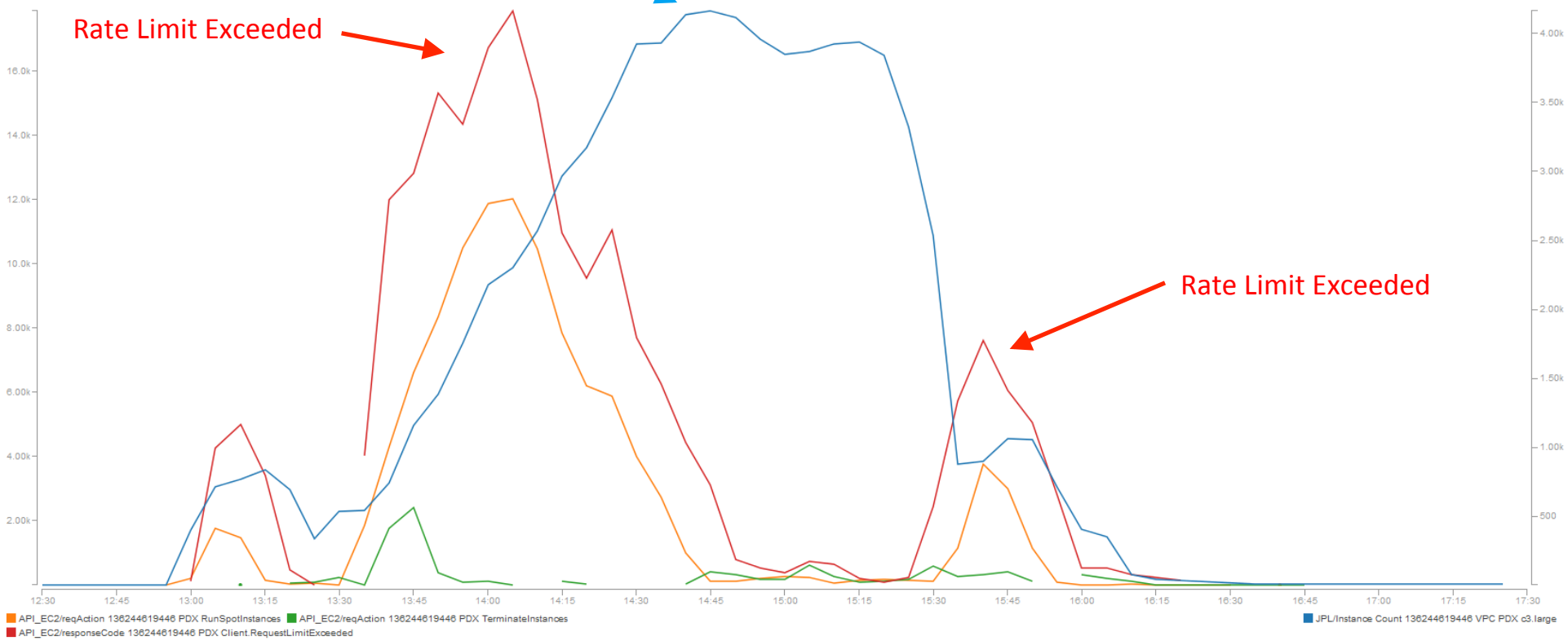
"Thundering Herd"

- Large fleet of auto-scaled compute instances calling same services at same time
  - *"API rate limit exceeded"*
- **"Jittering"** the API calls
  - Introduce ***randomizations*** to API calls
  - Distributes load on infrastructure



Large-Scaling API Rate Limits

- Example showing rate limits when scaling up to 4000+ EC2 compute instances and scaling down



SUMMARY

Key Points

- Cloud use can be cheaper if can dive deeper in architecture design and TCO impacts

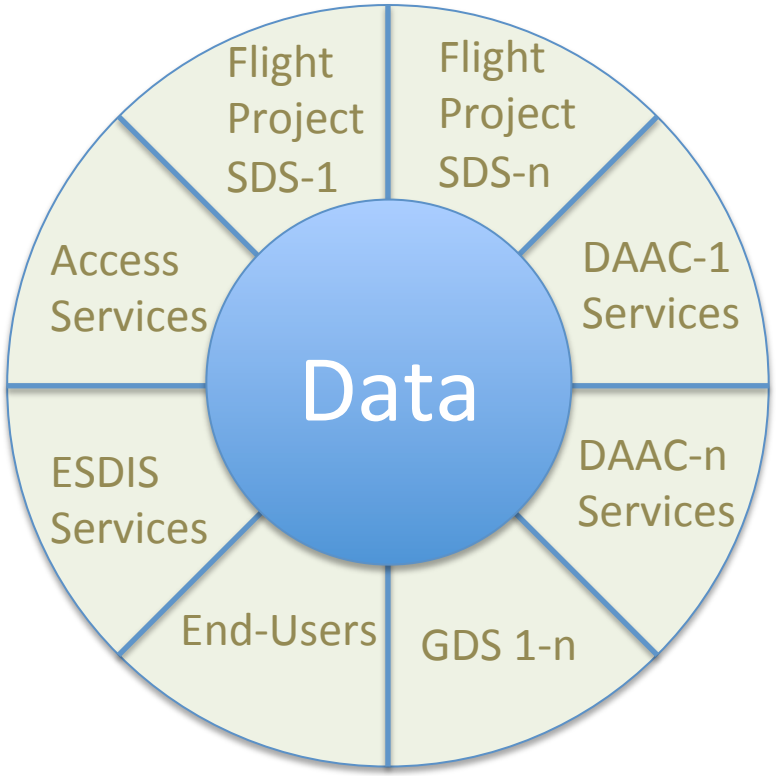
- Cost implications of Earth Science Data Systems in the Cloud:
  - Compute
  - Storage
  - Network
  - Deployment Topology
- Total Cost of Ownership (TCO) analysis
- At large scales, need to deal with **scaling issues**.
- Running on **spot market** for cost savings—but need **resiliency**
- Can achieve large-scale production while saving costs
- **Fault tolerant** science data systems can scale better in cloud computing environments
- **Collocation**
  - Data Lakes

- **Benchmark at full-scale!**
  - Test in full-scale production
  - Assess steady-state at full-scale
  - Monitor real-time metrics

"Data Lake"-extended

- It's about collocation!
- Minimize data movement
- Maximize user services
- Run on *public cloud provider* or at an *on-premise data center*

Reduce redundancy and foster collocated services



Enabling multi-disciplinary data approach for analysis

"Hot Data Lake"

- Long-term **public cloud storage is expensive** at PB-scales
  - …Unless we can negotiate deals with cloud vendors
- Use object store data lake for ***"hot data"***
  - SDSes generate data into object stores
  - Object stores contain "fresh" / "hot" data as rolling storage
  - Offline moving of older data to DAAC for permanent storage
  - Automate caching of "hot data" back from DAACs