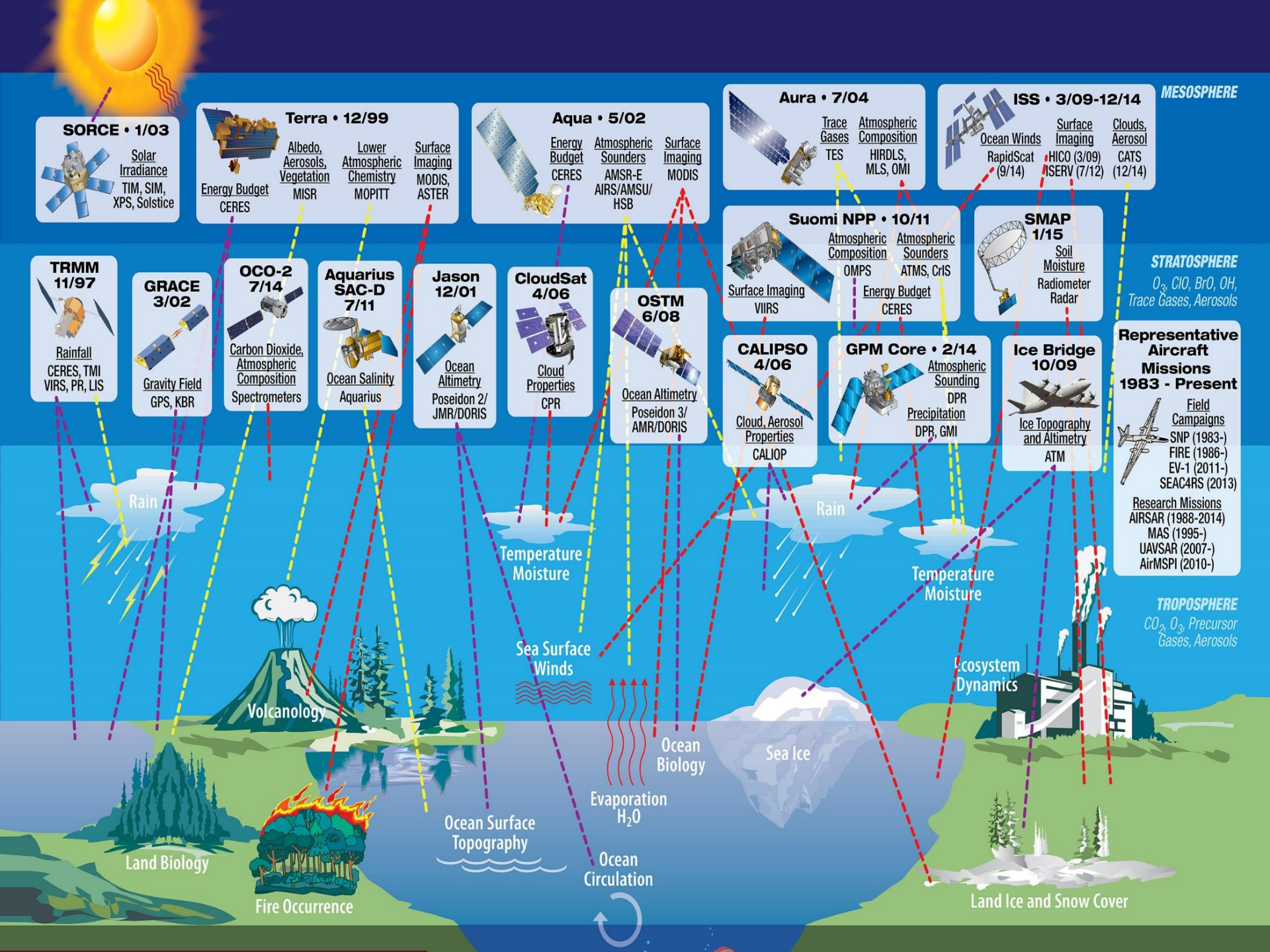
Lessons Learned while Exploring Cloud-Native Architectures for NASA EOSDIS Applications and Systems

Dan Pilone (dan@element84.com)

NASA EED2 Program

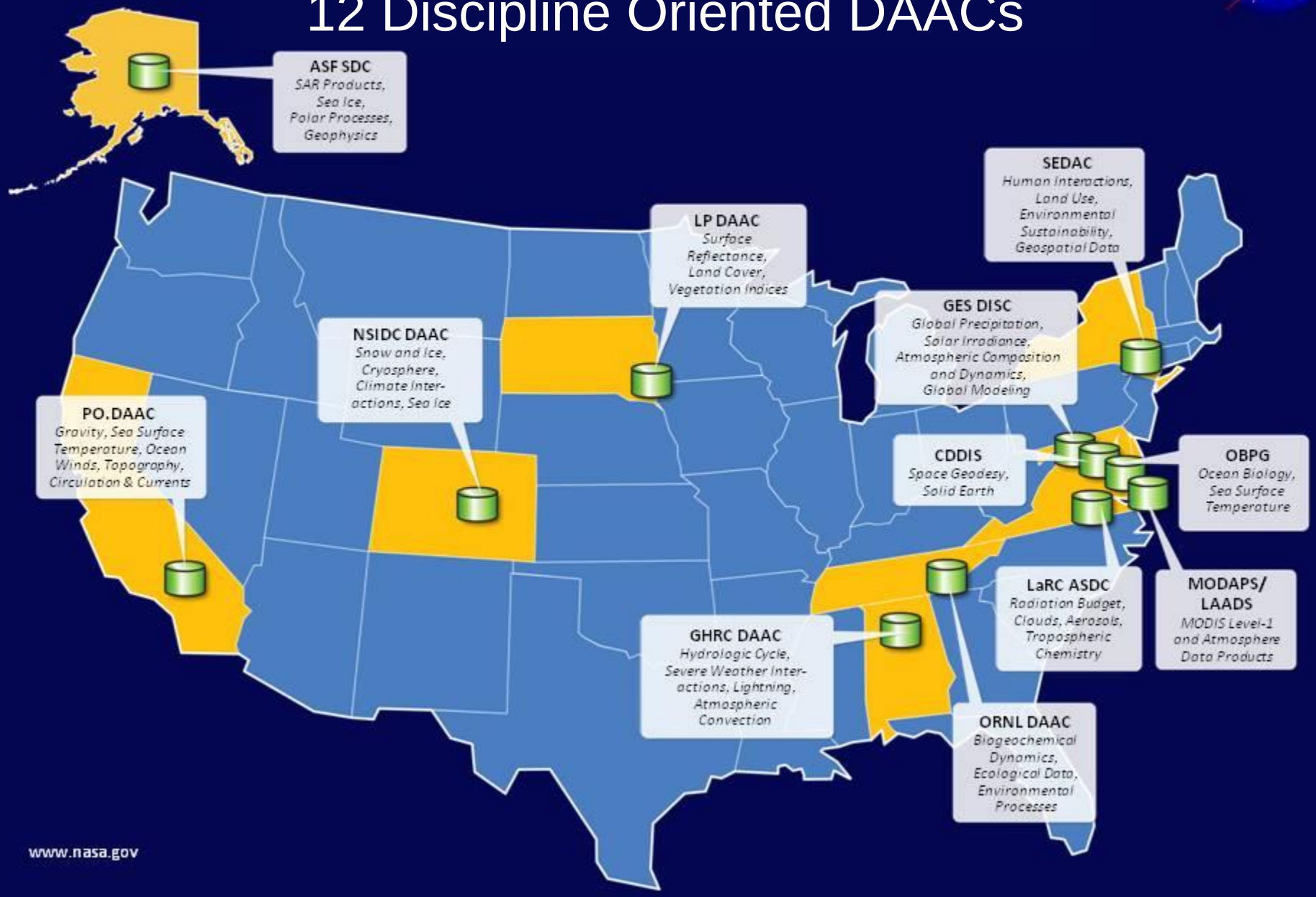
The material is based upon work supported by the National Aeronautics and Space Administration under Contract Number NNG15HZ39C


Formulation
Implementation
Primary Ops
Extended Ops
Sentinel-6A/B
Earth Science Instruments on ISS:
RapidScat, CATS,
LIS, SAGE III (on ISS), TSIS-1, OCO-3, ECOSTRESS,
GEDI, CLARREO-PF
Landsat 9
PACE
NI-SAR
SWOT
TEMPO
JPSS-2 (NOAA)
RBI, OMPS-Limb
GRACE-FO (2)
ICESat-2
CYGNSS
ISS
SORCE,
TCTE (NOAA)
NISTAR, EPIC
(NOAA'S DSCOVR)
QuikSCAT
EO-1
Landsat 7
(USGS)
SMAP
Terra
Suomi NPP (NOAA)
Aqua
CloudSat
Landsat 8
(USGS)
CALIPSO
Aura
GRACE (2)
OSTM/Jason 2
(NOAA)
GPM
OCO-2


SORCE • 1/03
Solar Irradiance
TIM, SIM, XPS, Solstice
Terra • 12/99
Energy Budget
CERES
Albedo, Aerosols, Vegetation
MISR
Lower Atmospheric Chemistry
MOPITT
Surface Imaging
MODIS, ASTER
Aqua • 5/02
Energy Budget
CERES
Atmospheric Sounders
AMSR-E AIRS/AMSU/ HSB
Surface Imaging
MODIS
Aura • 7/04
Trace Gases
TES
Atmospheric Composition
HIRDLS, MLS, OMI
ISS • 3/09-12/14
Ocean Winds
RapidScat (9/14)
Surface Imaging
HICO (3/09) ISERV (7/12)
Clouds, Aerosol
CATS (12/14)
MESOSPHERE
TRMM 11/97
Rainfall
CERES, TMI VIRS, PR, LIS
GRACE 3/02
Gravity Field
GPS, KBR
OCO-2 7/14
Carbon Dioxide, Atmospheric Composition
Spectrometers
Aquarius SAC-D 7/11
Ocean Salinity
Aquarius
Jason 12/01
Ocean Altimetry
Poseidon 2/ JMR/DORIS
CloudSat 4/06
Cloud Properties
CPR
OSTM 6/08
Ocean Altimetry
Poseidon 3/ AMR/DORIS
Suomi NPP • 10/11
Surface Imaging
VIIRS
Atmospheric Composition
OMPS
Atmospheric Sounders
ATMS, CrIS
Energy Budget
CERES
SMAP 1/15
Soil Moisture
Radiometer Radar
STRATOSPHERE
O_3, ClO, BrO, OH, Trace Gases, Aerosols
CALIPSO 4/06
Cloud, Aerosol Properties
CALIOP
GPM Core • 2/14
Atmospheric Sounding
DPR
Precipitation
DPR, GMI
Ice Bridge 10/09
Ice Topography and Altimetry
ATM
Representative Aircraft Missions 1983 - Present
Field Campaigns
SNP (1983-)
FIRE (1986-)
EV-1 (2011-)
SEAC4RS (2013)
Research Missions
AIRSAR (1988-2014)
MAS (1995-)
UAVSAR (2007-)
AirMSPI (2010-)
TROPOSPHERE
CO_2, O_3, Precursor Gases, Aerosols
Rain
Temperature
Moisture
Rain
Temperature
Moisture
Sea Surface Winds
Volcanology
Ecosystem Dynamics
Ocean Biology
Sea Ice
Evaporation H_2O
Land Biology
Ocean Surface Topography
Ocean Circulation
Fire Occurrence
Land Ice and Snow Cover

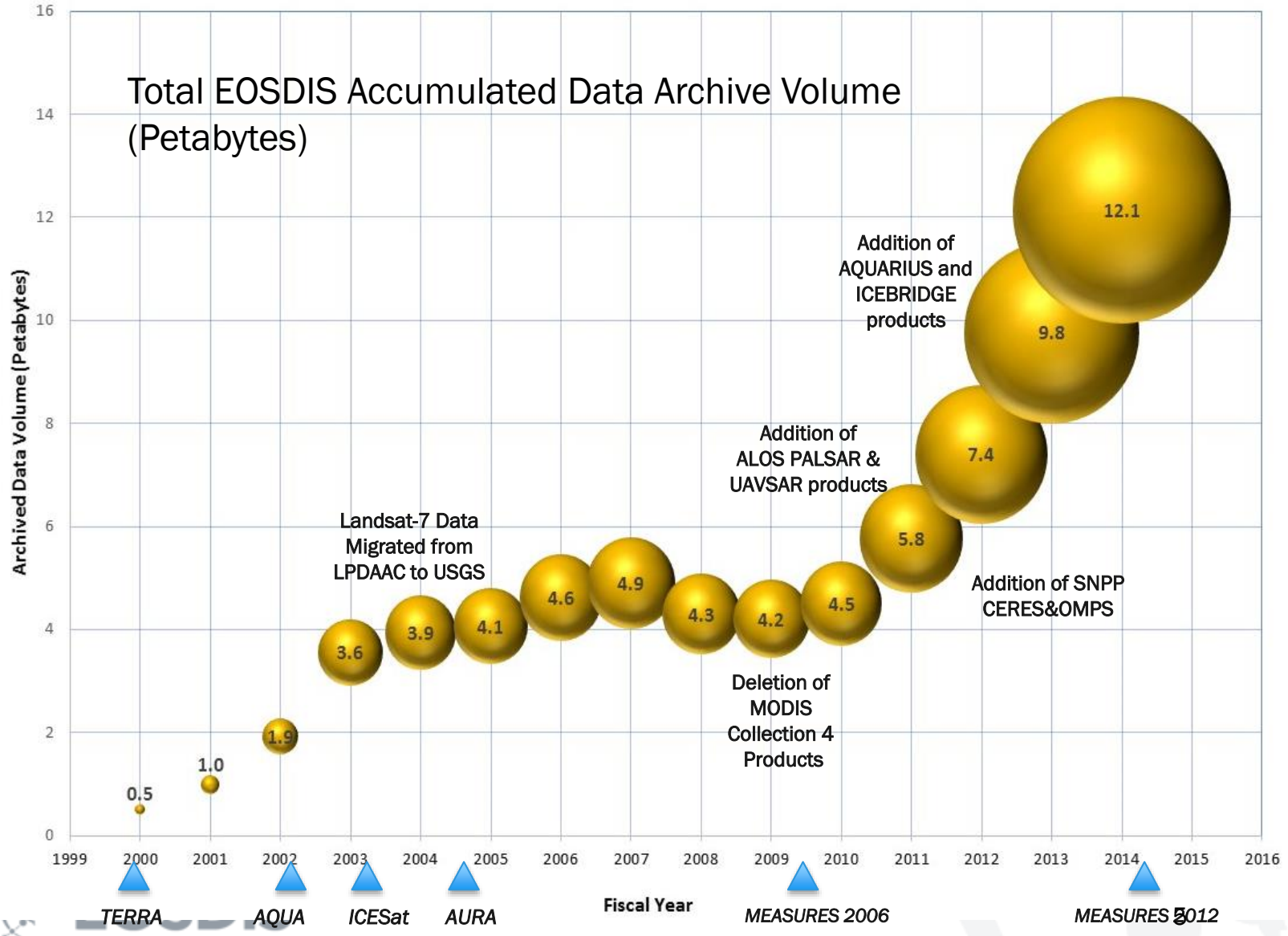
National Aeronautics and Space Administration

12 Discipline Oriented DAACs

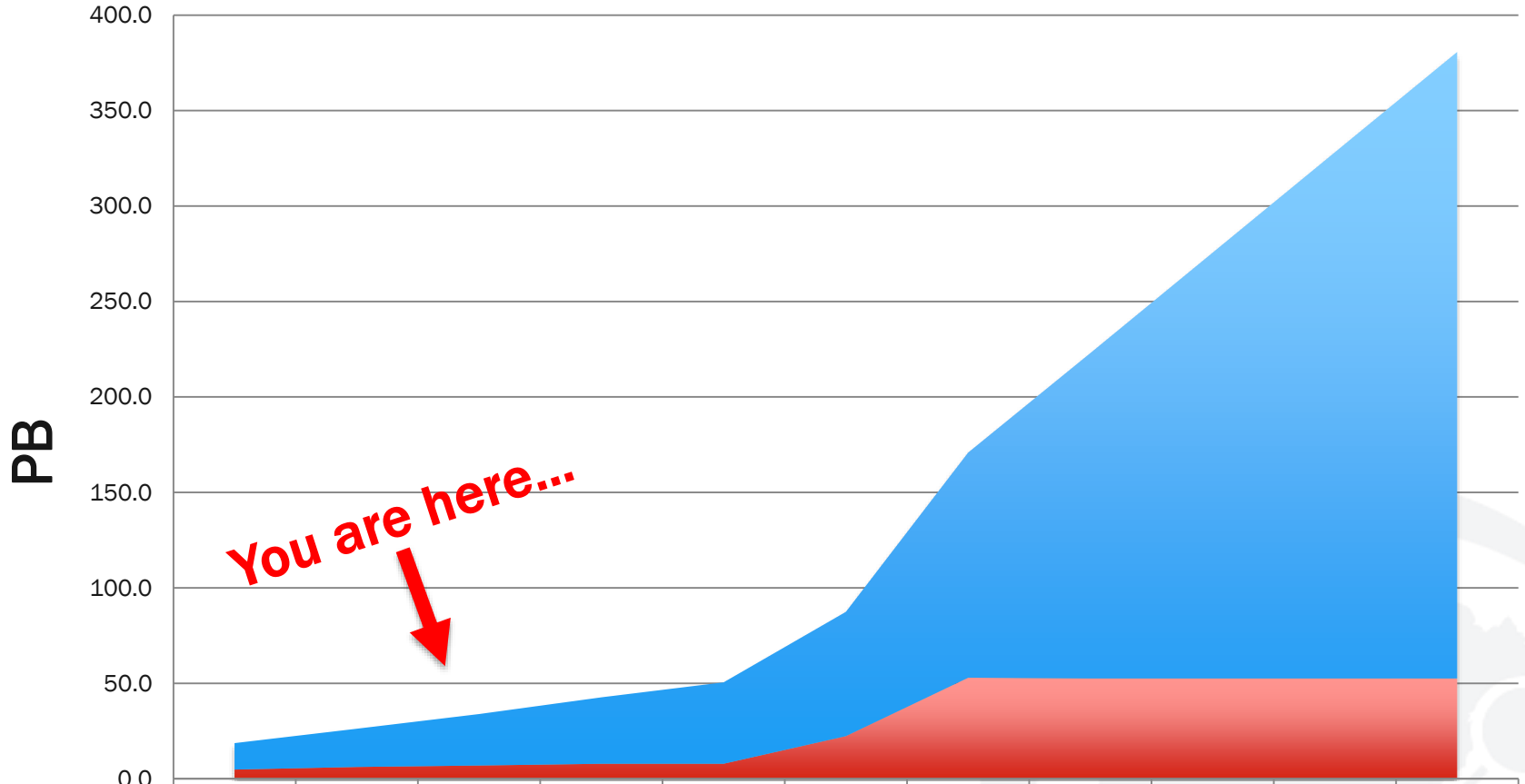


www.nasa.gov

15+ years of Earth Science Data



EOSDIS Archive Growth *Estimate* (Prime + Extended)



	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Cumulative Archive Size (PB)	13.8	20.0	27.0	34.8	42.7	65.0	118.0	170.5	223.1	275.6	328.2
Archive Growth Rate (PB)	4.9	6.2	7.0	7.9	7.9	22.4	52.9	52.6	52.6	52.6	52.6

EOSDIS

Lots of assumptions in this chart. Subject to change...

ExCEL Efforts and Project Prototypes

NGAP

NASA Compliant General Application Platform (NGAP), an operational, dev-ops, and sandbox AWS cloud based operating environment.

ASF WOS Prototype

AWS/NGAP Web Object Storage (WSO) prototyping large volumes of mission data dynamically between AWS S3, S3-IA, and Glacier object storage. Managed out of Alaska Satellite Facility

Earthdata Search Client to Cloud

NASA Earth Science data search by keyword and advanced filters such as time and space

Cumulus

Prototype addressing core EOSDIS capabilities including data ingest, archive, management, and distribution of large volumes of EOS data.

NISAR Preparation Prototype

Integrated prototype of science product generation and delivery from a DAAC system focused on coupling ASF DAAC and JPL ARIA systems.

CATEES

Easy-to-use Python tools packaged to support EOSDIS cross-DAAC science workflows and analytics over large volumes of EOS data in AWS.

ECC to Cloud Study

Earth Code Collaborative (ECC) study to determine cloud ready capabilities to migrate into AWS/NGAP platform.

ExCEL Efforts and Project Prototypes Continued

GIBS in the Cloud

Migrating GIBS to the AWS/NGAP Cloud based on recommendations made in the "GIBS in the Cloud Study"

Earthdata Login to Cloud Study

Study to determine and recommend migrating the Earthdata Login into AWS/NGAP cloud environment

CMR to Cloud

Migration of the Common Metadata Repository, into the AWS/NGAP platform based on recommendations made in the CMR to Cloud study.

OPeNDAP/HDF Cloud Studies

Study to determine and recommend a cloud native integration of OPeNDAP accessing HDF5 and netCDF4 data on AWS/NGAP platform.

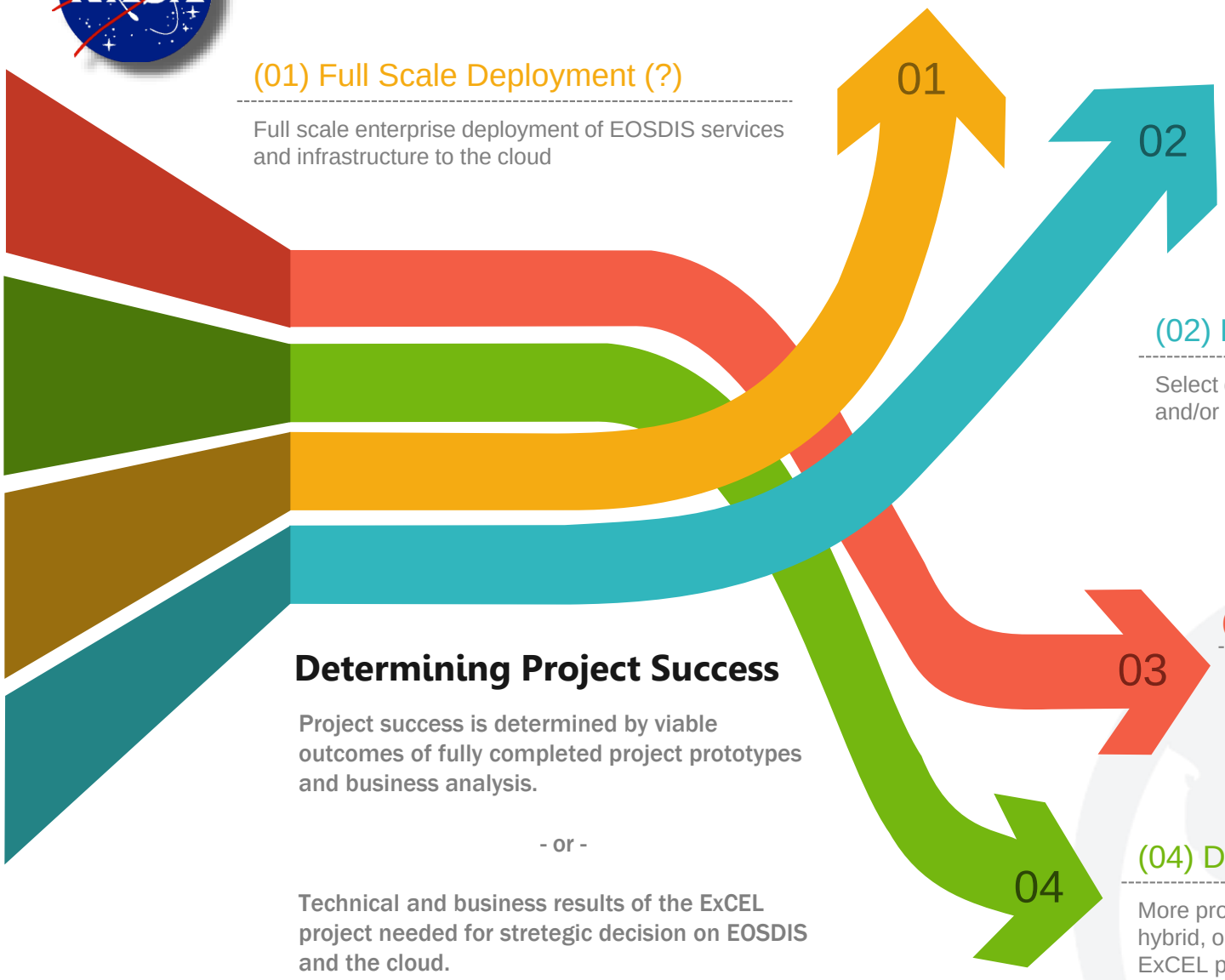
NEXUS

Prototype to accelerate end-user analysis of remote sensing data, highly parallel to better enable science discovery

Network Prototypes

Network prototypes to support to test security, monitoring, logging, and to perform R&D testing to support all ExCEL project prototypes.

ExCEL Go/No-Go



(01) Full Scale Deployment (?)

Full scale enterprise deployment of EOSDIS services and infrastructure to the cloud

(02) Partial Deployment (?)

Select deployment of EOSDIS services and/or infrastructure to the cloud

(03) Cloud Stand-down (?)

No EOSDIS services or infrastructure operationally migrated to the cloud

(04) Decision Point (?)

More prototyping required, or cloud hybrid, or other next steps based on ExCEL prototyping and business analysis results

Determining Project Success

Project success is determined by viable outcomes of fully completed project prototypes and business analysis.

- or -

Technical and business results of the ExCEL project needed for stretegic decision on EOSDIS and the cloud.

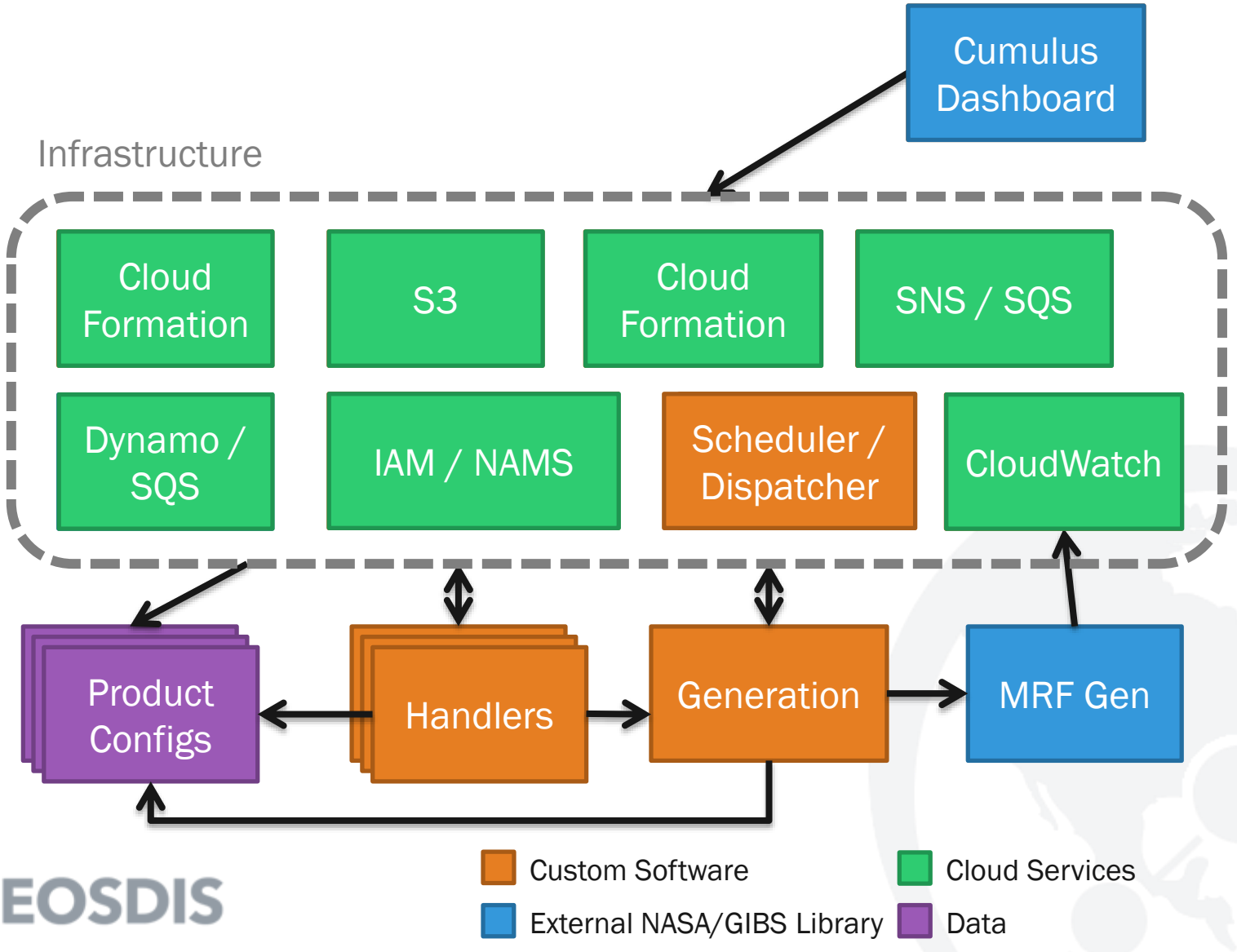
Lessons Learned

- Technical
- Cost
- Pyscho-Social

Technical Lesson 1

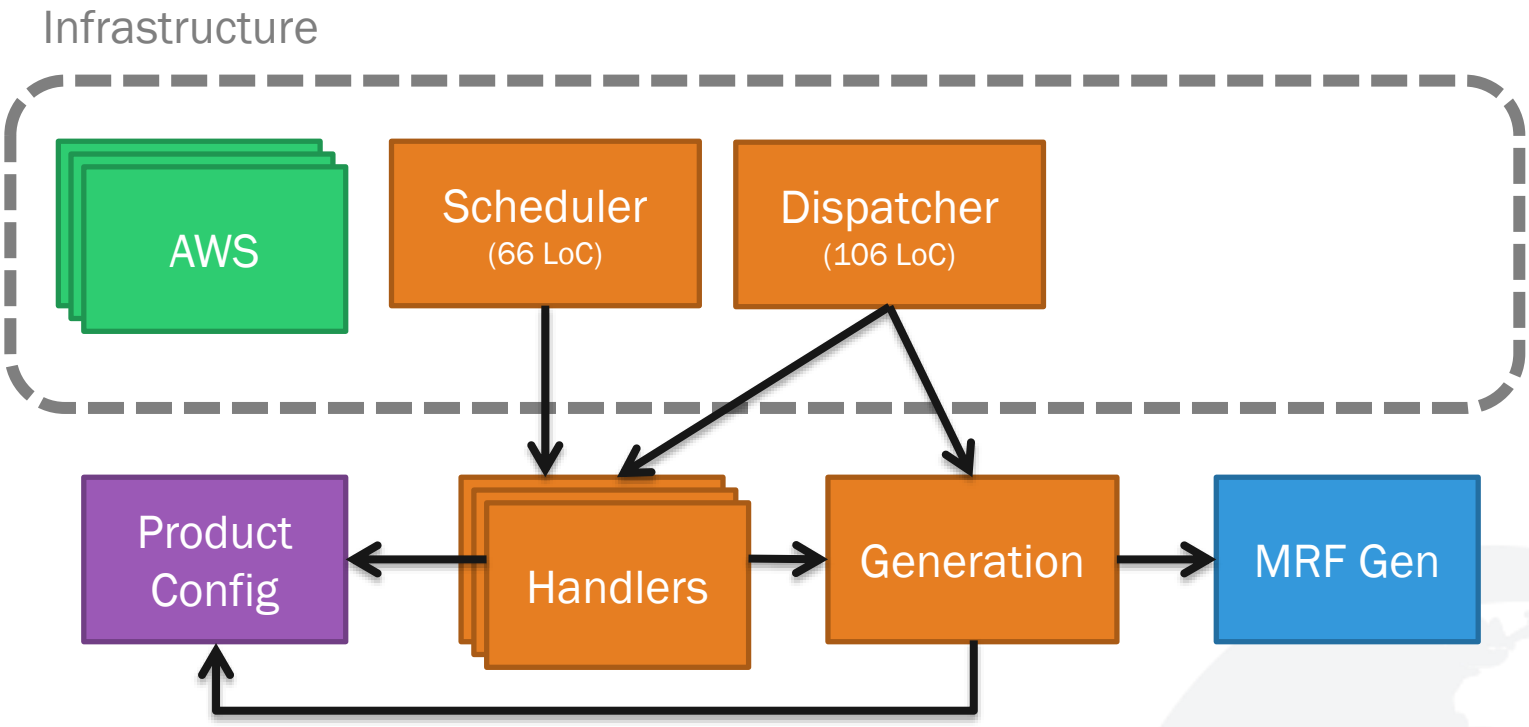
ENABLE CLOUD NATIVE ARCHITECTURES BY STRONGLY PREFERRING CLOUD SERVICES

GIBS-in-the-Cloud Service Swap



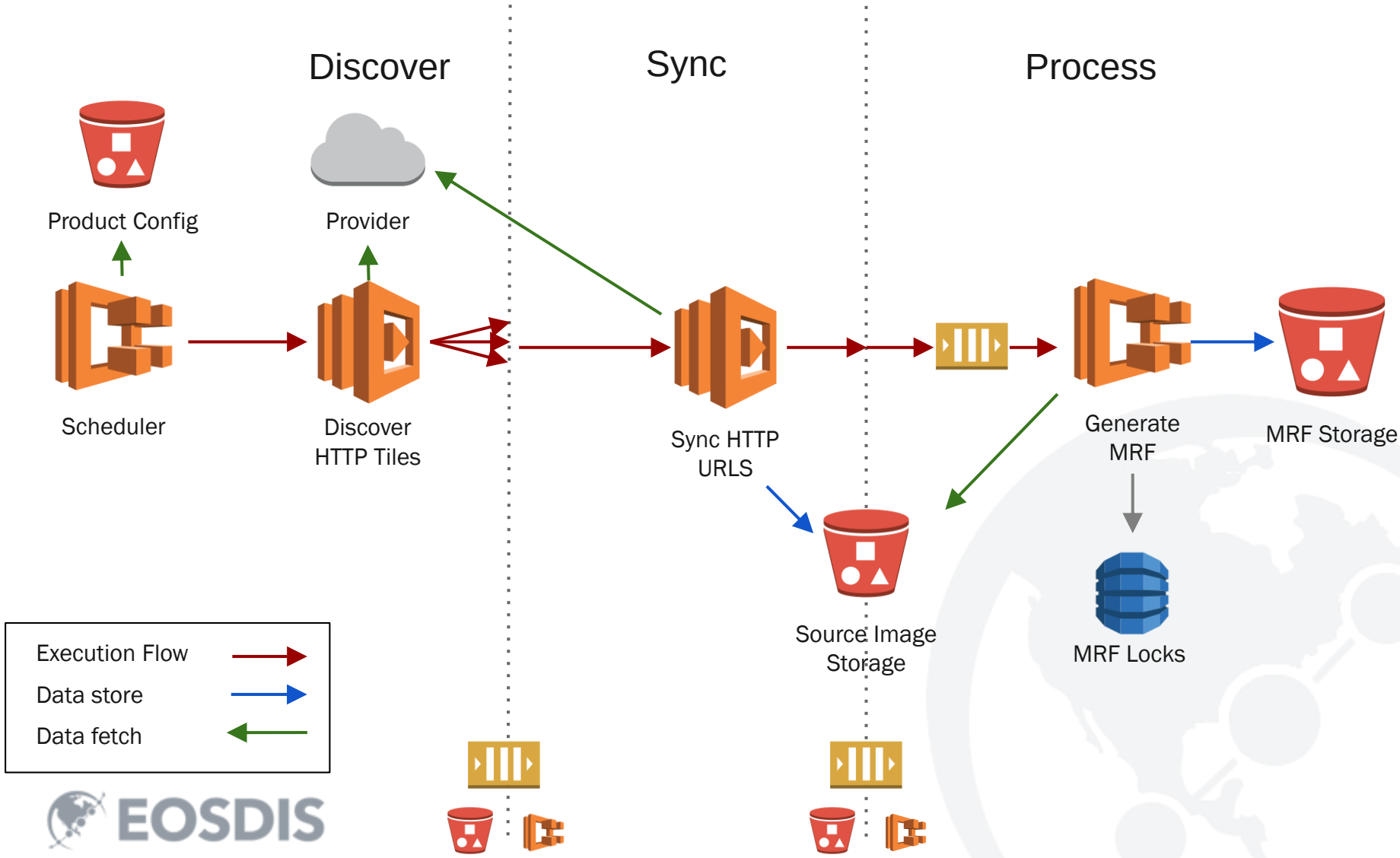
EOSDIS

GIBS-in-the-Cloud Ingest and Processing



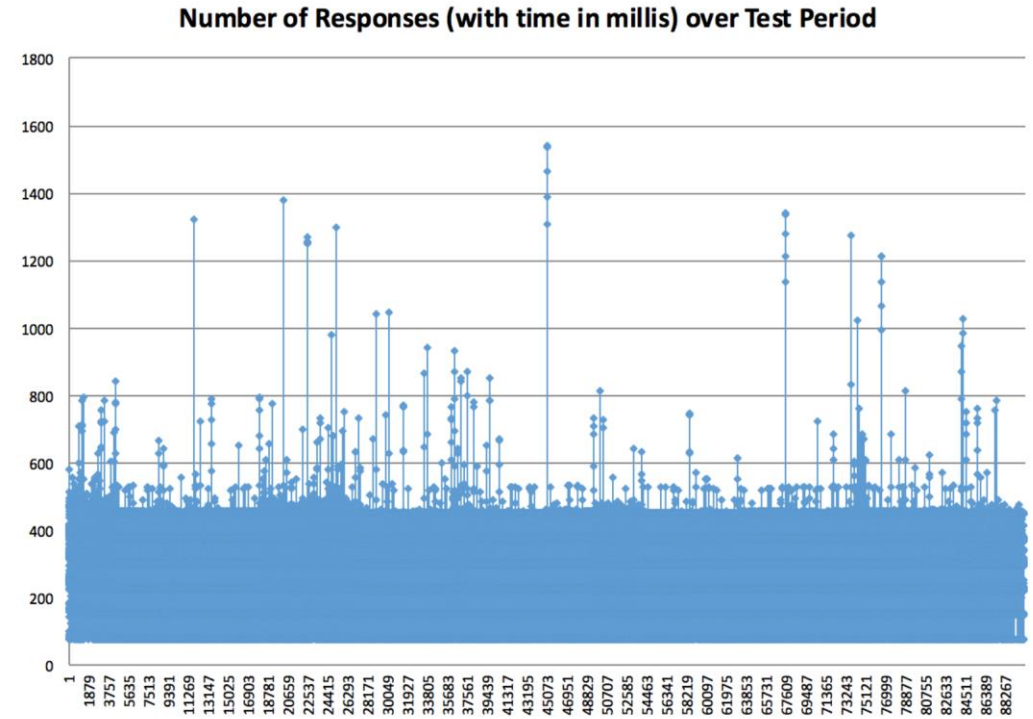
EOSDIS

Ingest: MODAPS Tiles

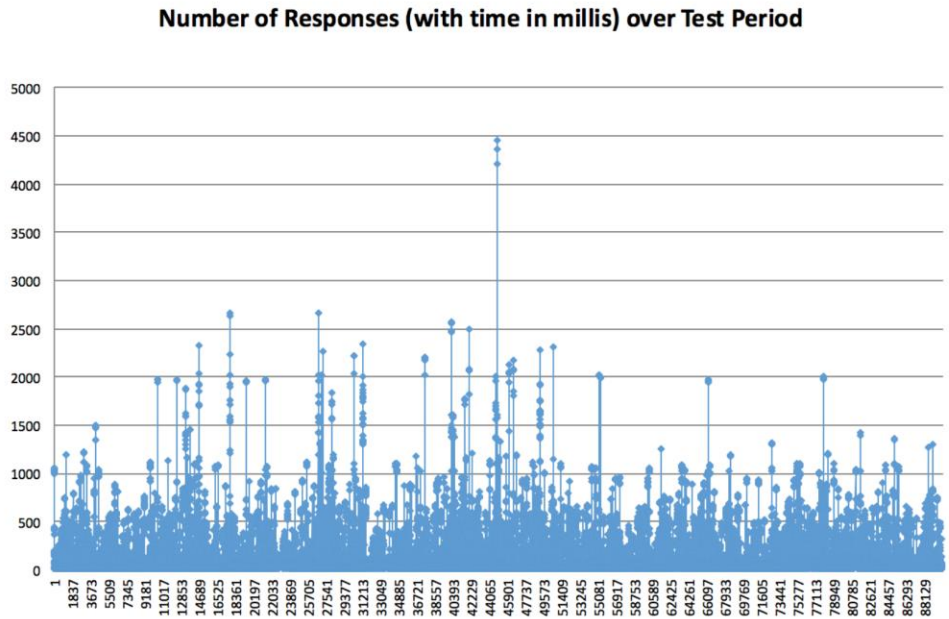


Technical Lesson 2

AWS HAS VERY LOW INTERNAL LATENCY – BUT TRUST NOTHING.



On premises implementation showed consistent performance during load testing vs more sporadic latencies in AWS.

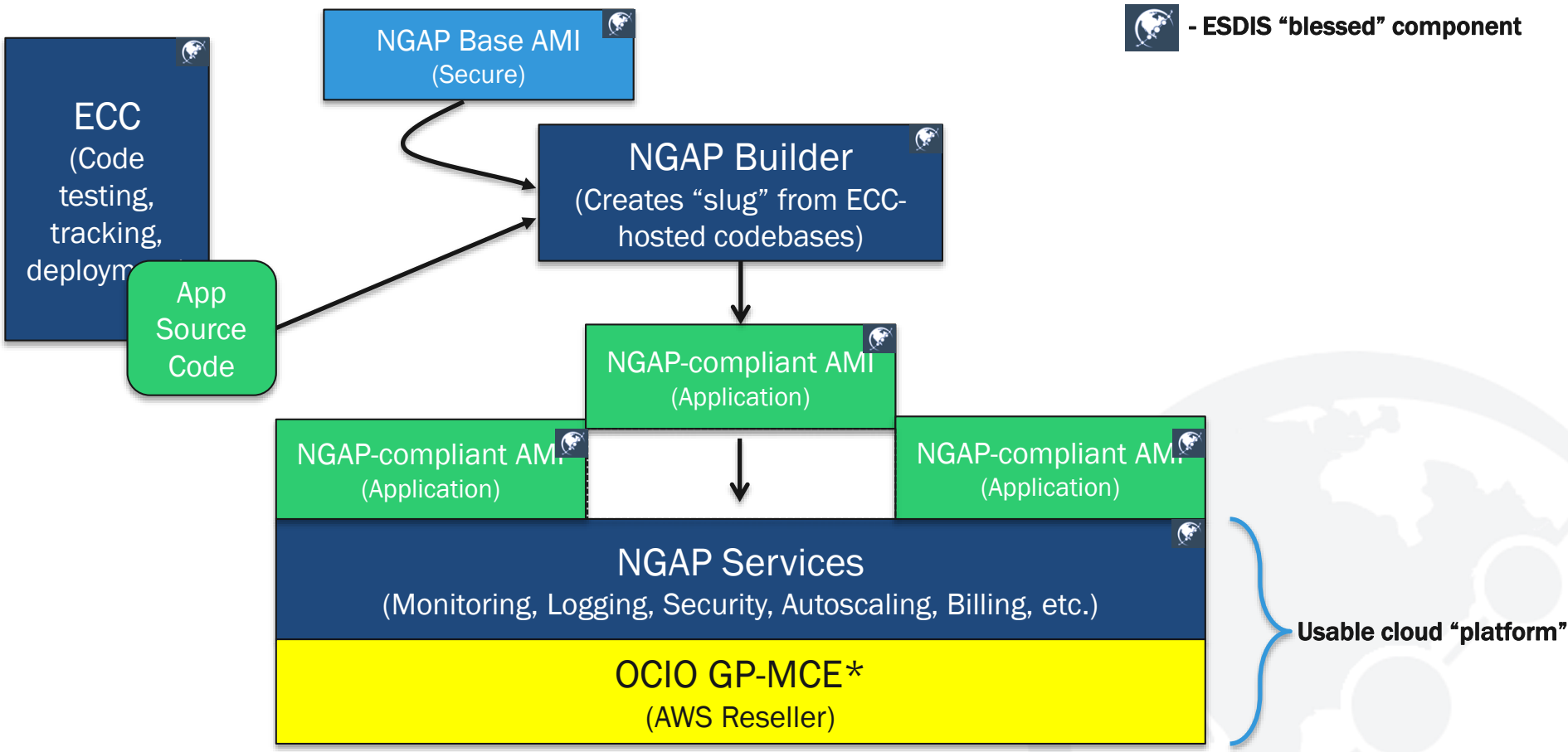


EOSDIS

Technical Lesson 3

INVOLVE SECURITY FROM THE VERY BEGINNING

Layer security thoughout the architecture



EOSDIS

*General Purpose Managed Compute Environment

Cost Lesson 1

MODELING TOTAL COST OF OWNERSHIP (TCO) IS EXTREMELY COMPLICATED

Need tear off edges here.

NGAP Application Profile	Web Application
Application Name	Earth Data Search
EC 2Instance Type	m4.xlarge; Tenancy: Shared; CPUs: 4; Memory: 16 GiB; Storage: EBS only: Cost: $0.239/Hrs
EC2 Instance OS Disk Size (GB)	8
EC2 Instance OS Disk type	General Purpose SSD
EC2 Disk Provisioned IOPS	30
EC2 Disk IO Rate (million)	10
EC2 Number of instances	2
EC2 Utilization Rate (% of day)	50%
RDS DB Engine (No BYOL)	PostgreSQL
RDS Deployment Option	Multi-AZ
RDS Instance Type	db.m1.xlarge; CPUs: 4; Memory: 15 GiB
RDS Disk Type	Provisioned IOP SSD
RDS Storage Capacity (GB)	500
RDS Provisioned IOPS	100
RDS IO Rate (million)	10
AWS Lambda Usage (million requests per month)	1.2
AWS Lambda Memory Requirement (GB)	0.5
AWS Lambda Average Duration per Request (Seconds)	1
S3 Storage (TB)	3358
S3 PUT, COPY, POST, or LIST Requests per month	1000000
S3 GET and all other Requests	1000000
Cloudfront Egress (GB)	TBD
Dynamo DB Writes per Second	12
Dynamo DB Reads per Second	12
Dynamo DB Storage (GB)	1
DynamoDB Stream Requests (per month)	5000000
SQS Requests per Month (Each <= 64kb, A 100kb message is 2 requests)	2000000
SNS Mobile Push Messages	2500000
SNS Email	8500
SNS HTTPS	1300000
SNS Egress	TBD
CloudWatch Dashboards	1
CloudWatch Detailed Monitoring EC2	1
CloudWatch Custom Metric	6
CloudWatch Alarms	20
CloudWatch API Requests	120000
CloudWatch Logs Ingested (GB)	3.2
CloudWatch Logs Archived (GB)	3.2
CloudWatch Egress	TBD
Environment Scaling	
Production Regions	2
UAT/Production Cost Ratio	1
SIT/Production Cost Ratio	1
DIT/Production Cost Ratio	0.5

EC2	Instance (hr)
EC2	Storage(GB)/M
EC2	IOPS per GB M
EC2	IO Rate (millio
RDS	Instance (hr)
RDS	Storage(GB)/M
RDS	IOPS per GB M
RDS	IO Rate (millio
Lambda	Million reques
Lambda	GB-Seconds
S3	
First 1 TB / month	
Next 49 TB / month	
Next 450 TB / month	
Next 500 TB / month	
Next 4000 TB / month	
Over 5000 TB / month	
S3 Non GET Requests	Requests per
S3 GET	Requests per
Cloudfront	Egress GB
DynamoDB	Writes/Hour
DynamoDB	Reads/Hour
Dyanmo DB	Storage(GB)/M
Dynamo DB	Stream Read R
SQS	per request
SNS	Mobile Messa
SNS	Email Messag
SNS	HTTP (m)
SNS Egress	
CloudWatch	Dashboard
CloudWatch	Detailed/insta
CloudWatch	Metric/month
CloudWatch	Alarm/month
CloudWatch	Requests (100
CloudWatch	GB
CloudWatch	GB
CloudWatch	Egress GB

Calculator | Cost Summary | NGAP Profiles | EC2 Data | RDS Data | Storage Data | Lambda Data | DynamoDB data | SQS Data | SNS Data | Cloudwatch Data | +

EOSDIS

November announcements as of the 7th…

Most Recent Announcements from AWS

Date	Announcement
Nov 03	Now run real-time stream processing at scale with Apache Flink on Amazon EMR
Nov 03	AWS CodePipeline Introduces AWS CloudFormation Deployment Action
Nov 03	Continuously Deliver Changes to AWS CloudFormation Stacks with AWS CodePipeline
Nov 03	Amazon Lumberyard Beta 1.6 now available, introduces Twitch Metastream and more
Nov 02	Amazon RDS for Oracle now supports 11g to 12c Major Version Upgrade
Nov 02	AWS Directory Service for Microsoft Active Directory (Enterprise Edition) is now available in the US East (Ohio) Region
Nov 02	Amazon CloudWatch Events is now available in the China (Beijing) AWS region
Nov 02	Amazon SES Now Provides Fine-Grained Email Sending Metrics
Nov 02	Backup to AWS with HP Data Protector 9.07 using Storage Gateway VTL
Nov 01	Swift Web Applications on the AWS Cloud: Quick Start Reference Deployment

EOSDIS

This is before considering…

- User behavior
- Staff cost savings
- Development cost savings
- Inter-region costs
- Data lifecycle modeling
- Application migration costs – both in and out
- Managing “consumption” based cost model

Cost Lesson 2

EXPLORE ALTERNATIVE ARCHITECTURES FOR POSSIBLE COST SAVINGS

Need graphic for multiple egress models, requestor pays architecture, etc.

Psycho-Social

GO HANDS-ON QUICKLY

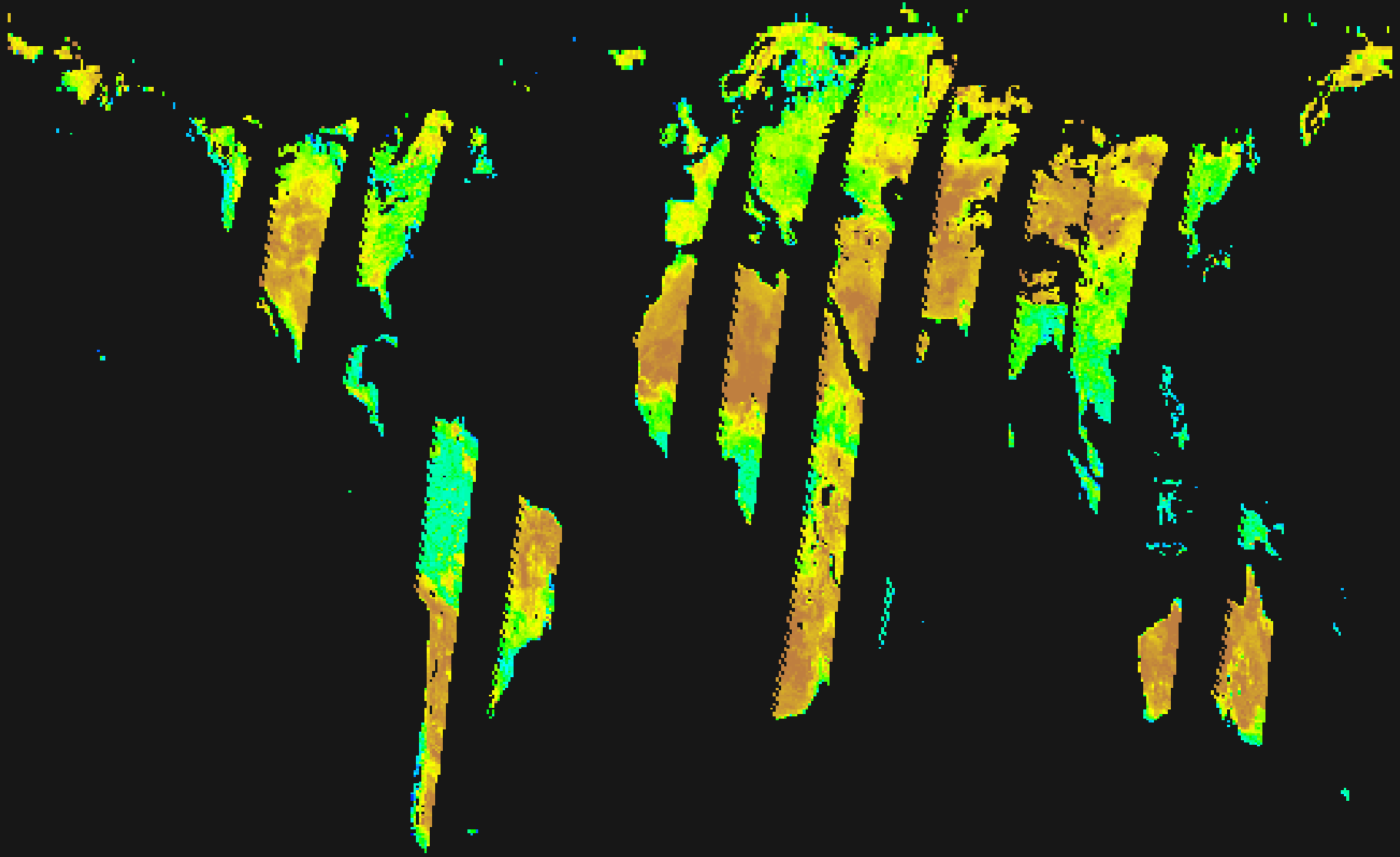

EED2
HACKFEST
2016
CAPABLE OF
SPENDING ENTIRE
ANNUAL BUDGET


EED2
HACKFEST
2016
MOST LIKELY TO
GO INTO
OPERATIONS


EED2
HACKFEST
2016
MOST
INNOVATIVE


EED2
HACKFEST
2016
MOST
POINTS

EOSDIS



Summary

- Enable cloud native architectures by strongly preferring cloud services
- AWS has very low internal latency, but trust nothing.
- Involve security from the very beginning
- Modeling TCO is extremely complicated
- Explore alternative architectures for possible cost savings
- Go hands-on quickly

Questions?

Dan Pilone

dan@element84.com

This material is based upon work supported by the National Aeronautics and Space Administration under Contract Number **NNG15HZ39C.**

Raytheon

EOSDIS