

Migration to Cloud and Path to Modernization for JPSS Data Production System

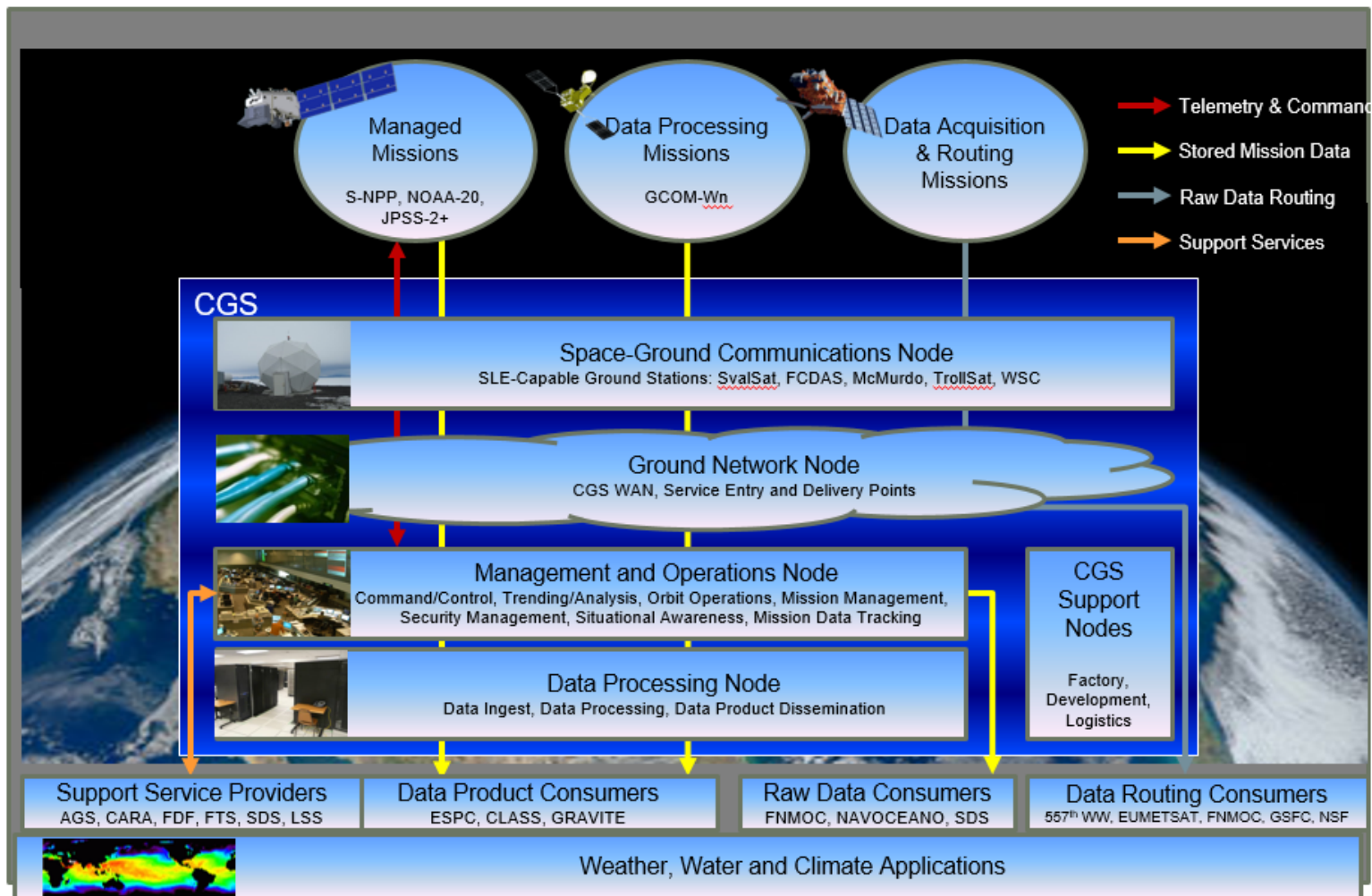
JPSS-CGS Raytheon IIS

Scott Kern

Josh Olson

Copyright © 2020, Raytheon Company. All rights reserved.
Published by The Aerospace Corporation with permission.

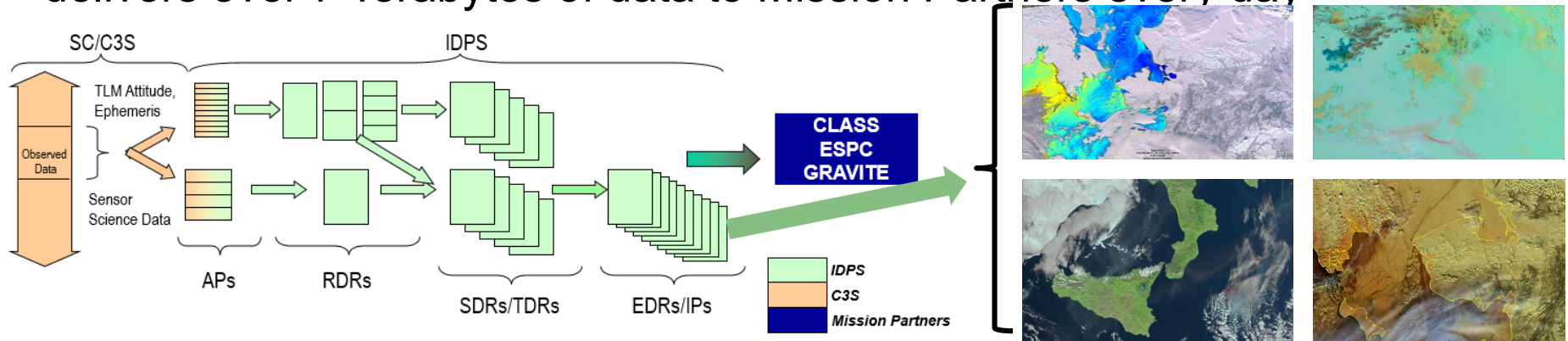
JPSS CGS Data Production – What is it?



DPN is the Joint Polar Satellite System - Common Ground System segment which provides ground data processing to create S-NPP and JPSS data products from raw sensor data.

JPSS CGS DPN – Details

- **Data Processing Node (DPN)**
 - Ingests Mission Data packets (APs) received from C3S
 - Produces Data Products: RDRs, SDRs, TDRs, EDRs, IPs
 - Delivers to Mission Partners: CLASS, GRAVITE, and ESPC
- **Key Architectural Features**
 - Configurable data driven algorithm processing chains
 - Data is processed for the S-NPP, N20 and GCOM-W missions
 - IDP deployed at NSOF and CBU – moving to the Cloud
 - Processing load balanced for fault management
 - NIST 800-53 v3 security implementation
 - Focus on Low Data Latency and High Availability of data products
- **DPN operations receives ~400 Gigabytes of data from 3 spacecraft and delivers over 7 Terabytes of data to Mission Partners every day**



DPN processes observed photons into deliverable products
Observation Time to MP Delivery: <100 min

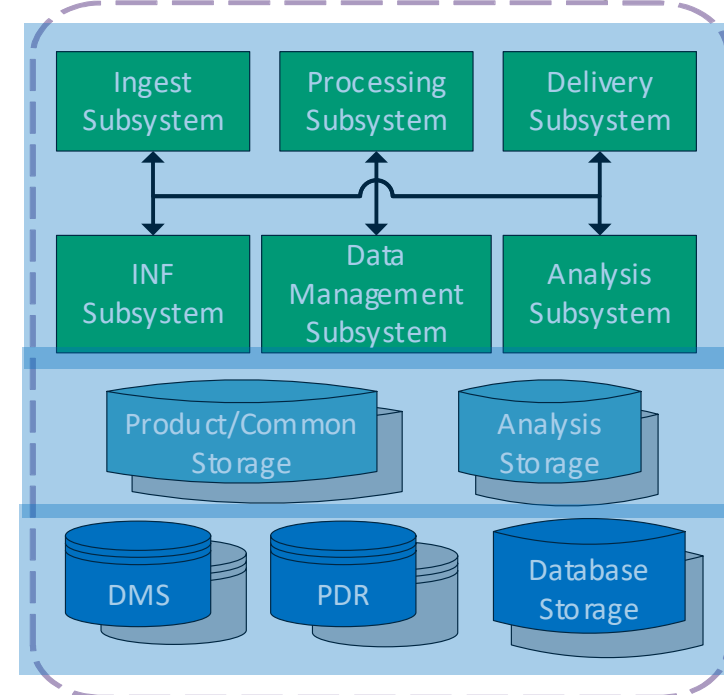
JPSS CGS DPN Cloud Migration – History and Milestones

- 2017: NASA/NOAA asked RTN to begin evaluating migrating to cloud and produce whitepapers exploring viability
- 2018 Part 1: Proof-of-Concept deployment to AWS
 - Initial estimate was ~1 month to get DPN running after environment configuration
 - **Completed in ~4 days!**
- 2018 Part 2: RTN DevCloud Prototype/Demos
 - Execute Trade Studies and evaluate end-to-end system performance with security tools in place
 - Demonstrated cross-AZ failover using AWS RDS DBaaS
- 2018 Part 3: MS Azure
 - Risk reduction to ensure DPN deployment to MS Azure
 - NOAA Cloud Initiative was indicating preference for Azure
- 2019:
 - NOAA determined that DPN will be one of the first major NOAA programs to migrate to AWS GovCloud
 - Includes dedicated environments for:
 - Operations
 - Integration and Test
 - Factory/Development
 - Algorithm Development and Assessment
- Design Review completed January 2020
- **Transition to Operations December 2020**

Initial Implementation – Phase 1

- Transition to Operations in Cloud must occur NLT EOY 2020 (Lenovo HW waiver expiration)
- NOAA direction to migrate current operational baseline to Cloud with minimal baseline changes
 - Only changes to baseline that are explicitly necessary to operate in the cloud
 - Migrate primary DB from Oracle to PostgreSQL to save licensing costs
- HOT backup of primary Operations DP
 - Security Patching requires transition to backup IDP
 - 3rd IDP necessary to accommodate monthly patches and baseline upgrades while maintaining resiliency to failures
- Primary change is new Common Environment :
 - Route data to multiple DPN systems from a single on-prem data source
 - Management of security functions
- Leveraging DevOps Tools/Processes:
 - Environments 100% managed using Infrastructure-as-Code (Packer, Terraform, Chef)
 - Faster/Frequent algorithm releases to PRO subsystem decreases Research-to-OPS (R2O) cycle
- ~60 EC2 VMs and 500 TB storage per DP

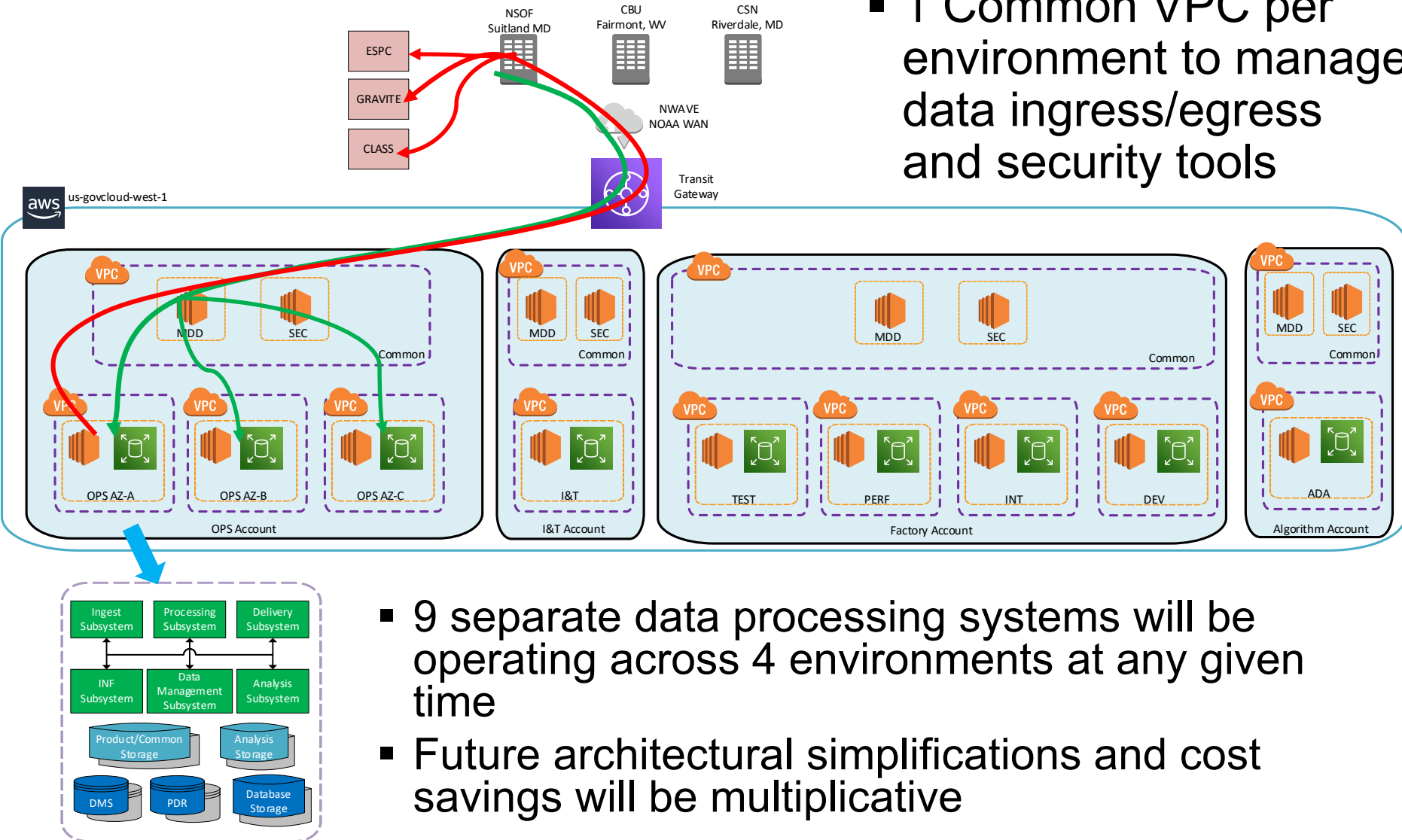
Operational Data Processor Configuration



- Database Layer (EC2 and EBS)
- Oracle Dataguard installed to EC2
 - Backup DB instance
 - EBS storage attached to EC2
 - DMS: Data Management
 - PDR: Performance Data Repo

Initial Implementation – Phase 1

- 1 Common VPC per environment to manage data ingress/egress and security tools



- 9 separate data processing systems will be operating across 4 environments at any given time
- Future architectural simplifications and cost savings will be multiplicative

Optimization – Phase 2

- Optimization Phase Updates the DPN cloud design to take better advantage of cloud capabilities
- Provides significant cost savings over initial-implementation
 - Savings for Infrastructure, COTS, O&M
- Implements a better foundation for science/forecast product driven changes during Modernization Phase

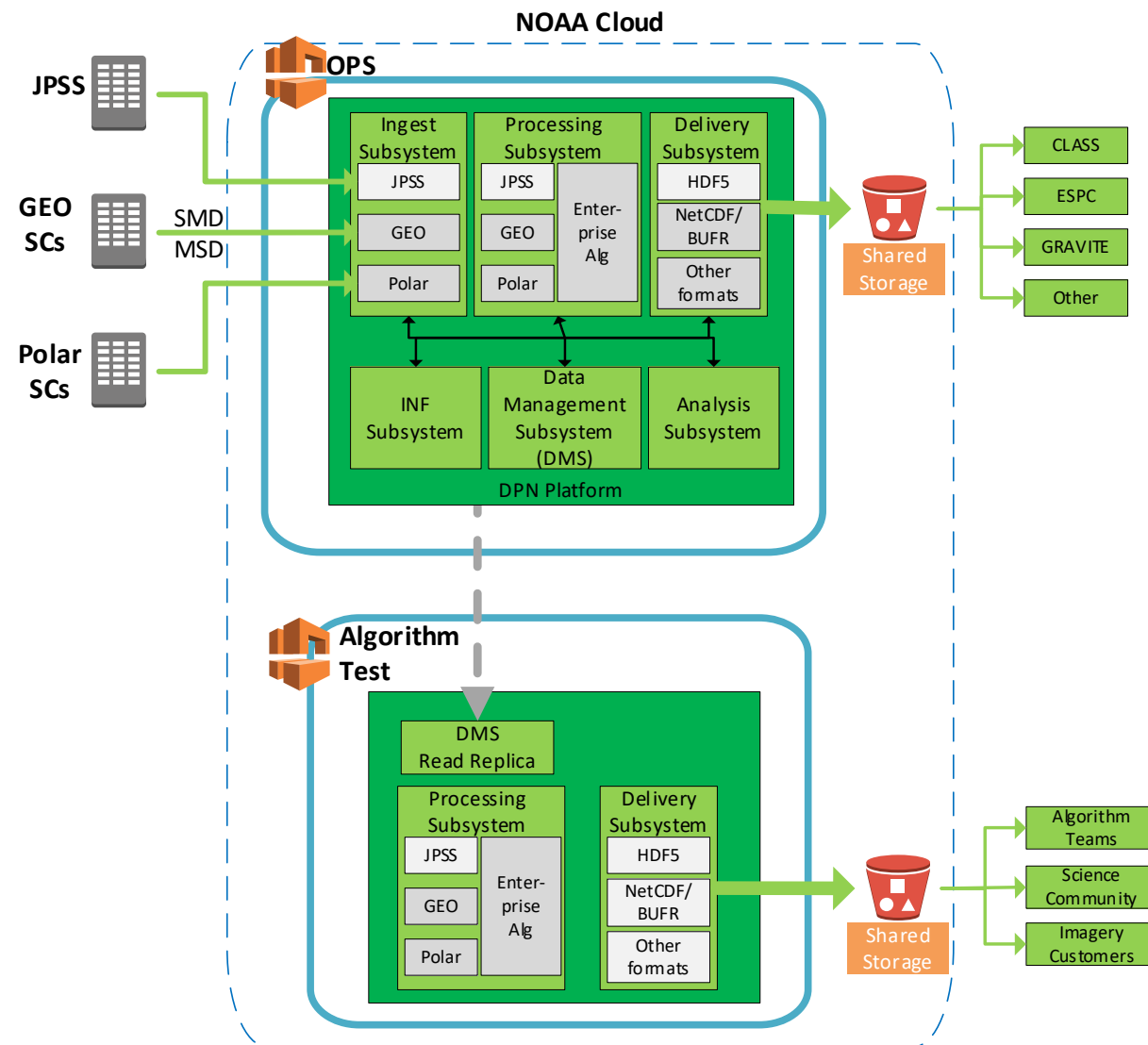
Optimization	Description
Transition to Highly Available (HA) DPN	<ul style="list-style-type: none"> • Deploy single HA DP spanning 2 Availability Zones <ul style="list-style-type: none"> • Subsystems deployed across AZs in auto-scaling groups • “Live” security patching on dynamic instances to eliminate OPS/Non-OPS transitions for monthly security patching
Dynamic Allocation of Processing Capacity	<ul style="list-style-type: none"> • Elastic processing capacity to dynamically respond to changing throughput needs in responding to anomalies
Complete migration of all databases to PostgreSQL	<ul style="list-style-type: none"> • COTS licensing savings • Reduces DBA support needs and security patching overhead
Modernize DPN Storage Layer	<ul style="list-style-type: none"> • Product storage moved from GPFS to cloud-native blob storage (AWS S3) <ul style="list-style-type: none"> • Significant cost savings • Initial prototyping shows satisfactory performance with minimal code modifications • Common storage migrates to cloud-native shared file system (AWS Elastic File Service EFS) <ul style="list-style-type: none"> • Provides HA without overhead required to manage large replicated storage cluster
Utilize Clustered Messaging Service	<ul style="list-style-type: none"> • Develop HA messaging system or utilize “Messaging-as-a-Service from AWS (Amazon MQ)
Utilize Cloud-Native Monitoring and Alerting	<ul style="list-style-type: none"> • Initial-Implementation using legacy design of monitoring agents deployed on DPN VMs delivering messages to operations.

Modernization – Phase 3

- Potential capabilities based on having DPN in the Cloud
 - The modernization phase could leverage DPN proven data production platform
 - Provide an expanded number of enterprise data products
 - Decreases algorithm process overhead accelerating R2O cycle
 - Data Delivery capability to expanded user base while minimizing data egress costs
 - Prioritize Real-time products critical to NWP delivered with DPN proven low-latency and stability
 - Products are packaged and delivered as needed with all products available in S3

Optimization	Description
Modernize Processing Subsystem using Containerized Algorithms	<ul style="list-style-type: none"> Science teams directly develop algorithms using containerized ADL and include dependencies in versioned containers Run multiple algorithm versions in parallel, dependencies reside in container Enterprise data product generation Real-time Processing: Operational algorithms generating products Off-line Processing: “Algorithm Sandbox” Evaluate updates to algorithms <ul style="list-style-type: none"> Executed during “back-orbits”, spot-instances or serverless Eliminates need for full DP dedicated for dedicated I&T and provides faster R2O cycles
Modernize Data Delivery via Cloud-based Content Delivery Network	<ul style="list-style-type: none"> Data products delivered to single cloud location (S3) <ul style="list-style-type: none"> Eliminate delivery of products through C3S facility to Mission Partners Real-Time Delivery: Products delivered to S3 location <ul style="list-style-type: none"> NWP products delivered in directly ingestible format (HDF, BUFR, NetCDR, etc) Consumers who need real-time products will receive notification of new products and API to pull the data directly down to their system (S3 => SNS => SQS pipeline) Off-Line Delivery: <ul style="list-style-type: none"> Non-Real-Time consumers will be able to request aggregation and/or packaging of products which will create a new product in S3 and notification delivered to consumer
“Lights Out” DPN decreases reliance on dedicated operations staff	<ul style="list-style-type: none"> DPN is highly stable system requiring almost no human interaction to function <ul style="list-style-type: none"> Decreases reliance on 24x7 dedicated operators Remove Java based GUIs and replace with simplified web GUI with APIs to drive DPN functions <ul style="list-style-type: none"> Significantly improves security posture

DPN Updates for Enterprise Algorithm Processing



- Production of new GEO/Polar mission products in parallel with JPSS OPS
 - Delivery to consumers in any format from data lake
- Potential to produce enterprise algorithms
- “New” algorithms can be executed in parallel to operations with no impact to OPS
 - Data Driven process leveraging database/storage read-replica

Summary

- JPSS CGS DPN team has outlined a path to modernization for the legacy processing system
- Some optimizations have already been prototyped and demonstrated
- Modernization Objectives:
 - Drive cloud operation costs down
 - Provide better turn around time for science and NWP products