SOLERS
A Peraton Company

Amazon Web Services (AWS) Cloud Platform for Satellite Data Processing

Small Sat 2019
SSC19-IX-01
Rich Baker
Chief Architect
Solers, Inc.
Email: richard.baker@solers.com
Phone: 240-790-3338

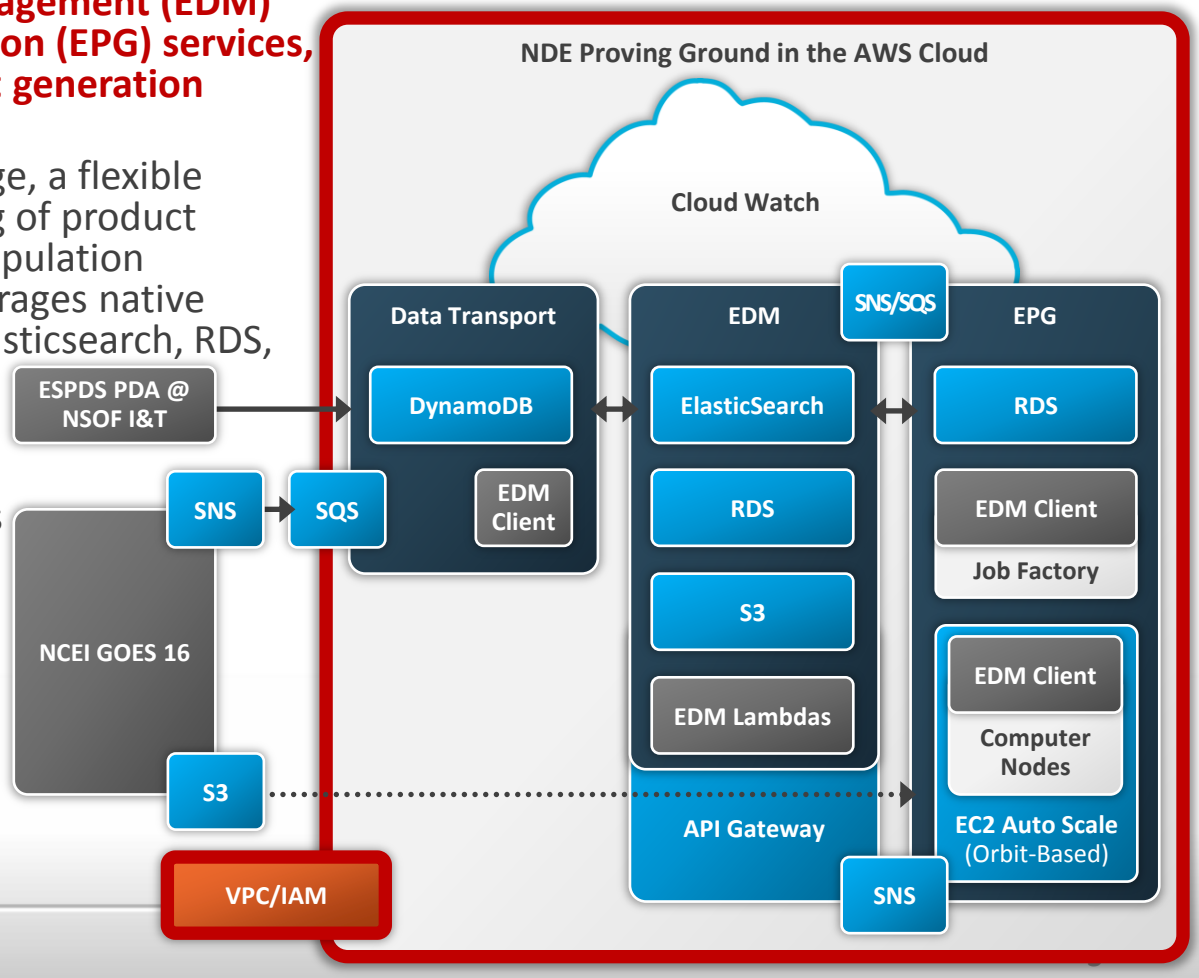
John Sobanski, Peter MacHarrie, Steve Causey, George Wilkinson, Steve Walsh, Ron Niemann, Dan Beall
Solers, Inc.

Agenda

- Overview of Solers' NOAA Data Exploitation (NDE) Proving Ground in the AWS Cloud
- Overview of AWS Ground Station
- Combining These Two to Create an Integrated AWS Cloud Platform for Satellite Data Processing
  - ***Potential use case / applicability to Small Satellites***

NOAA Data Exploitation (NDE) Proving Ground Overview

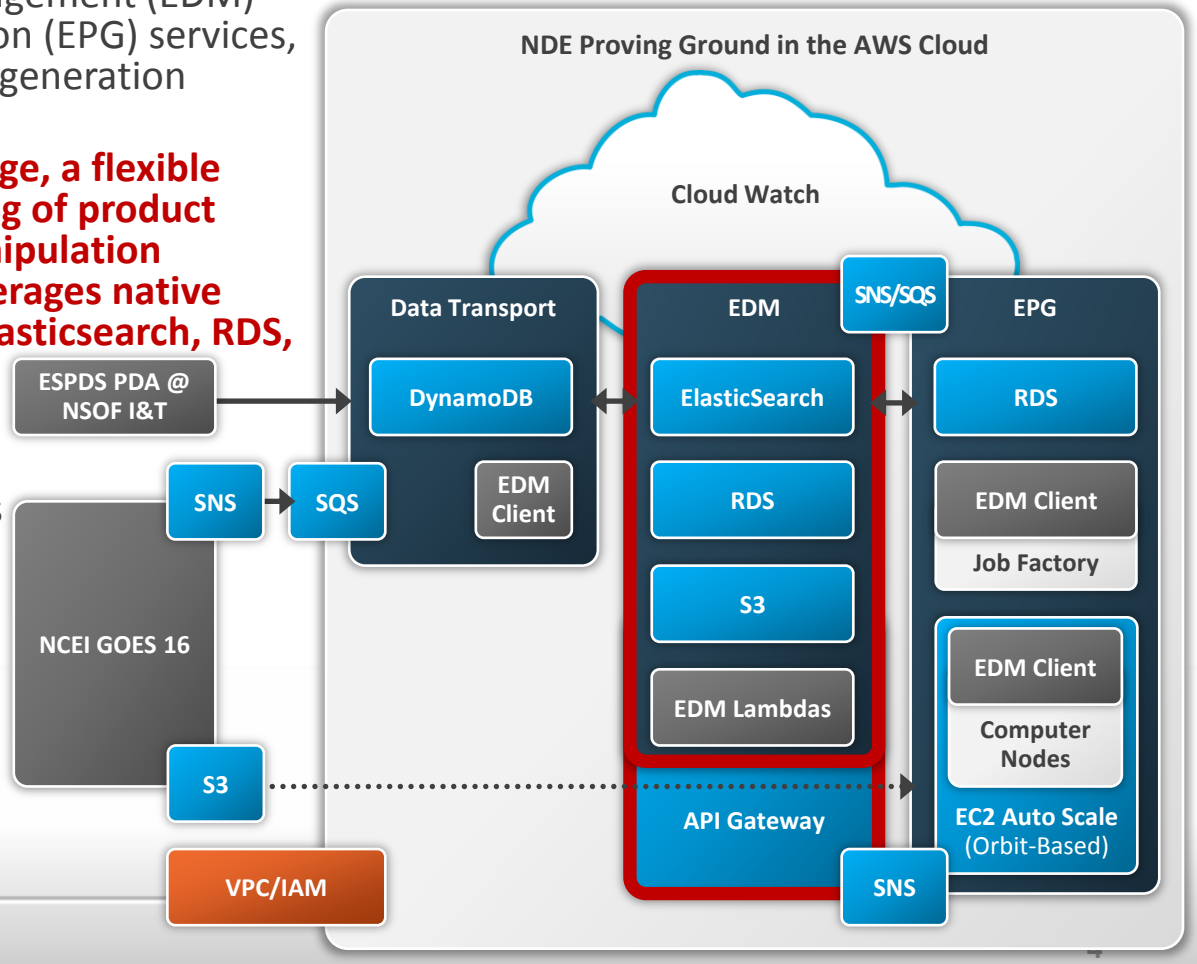
- **As part of NOAA's Environmental Satellite Processing and Distribution System (ESPDS) program, Solers created a cloud platform for satellite data management and processing, known as the NOAA Data Exploitation (NDE) Proving Ground.**
- **Re-factoring of NOAA's on-premise operational NDE system to use AWS cloud services.**
- **Consists of Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services, with a subset of NOAA's product generation algorithms integrated.**
- EDM service provides data storage, a flexible and searchable inventory/catalog of product metadata, and science data manipulation through RESTful interfaces. Leverages native AWS cloud services including: Elasticsearch, RDS, S3, Lambda, and API Gateway.
- EPG is capable of generating Level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.
- Data ingested:
  - GOES-16 data from the NOAA/NCEI Big Data Project (AWS S3 bucket).
  - S-NPP, JPSS-1, and GCOM-W data from ESPDS PDA at NSOF I&T.



SOLERS
A Peraton Company

NOAA Data Exploitation (NDE) Proving Ground Overview

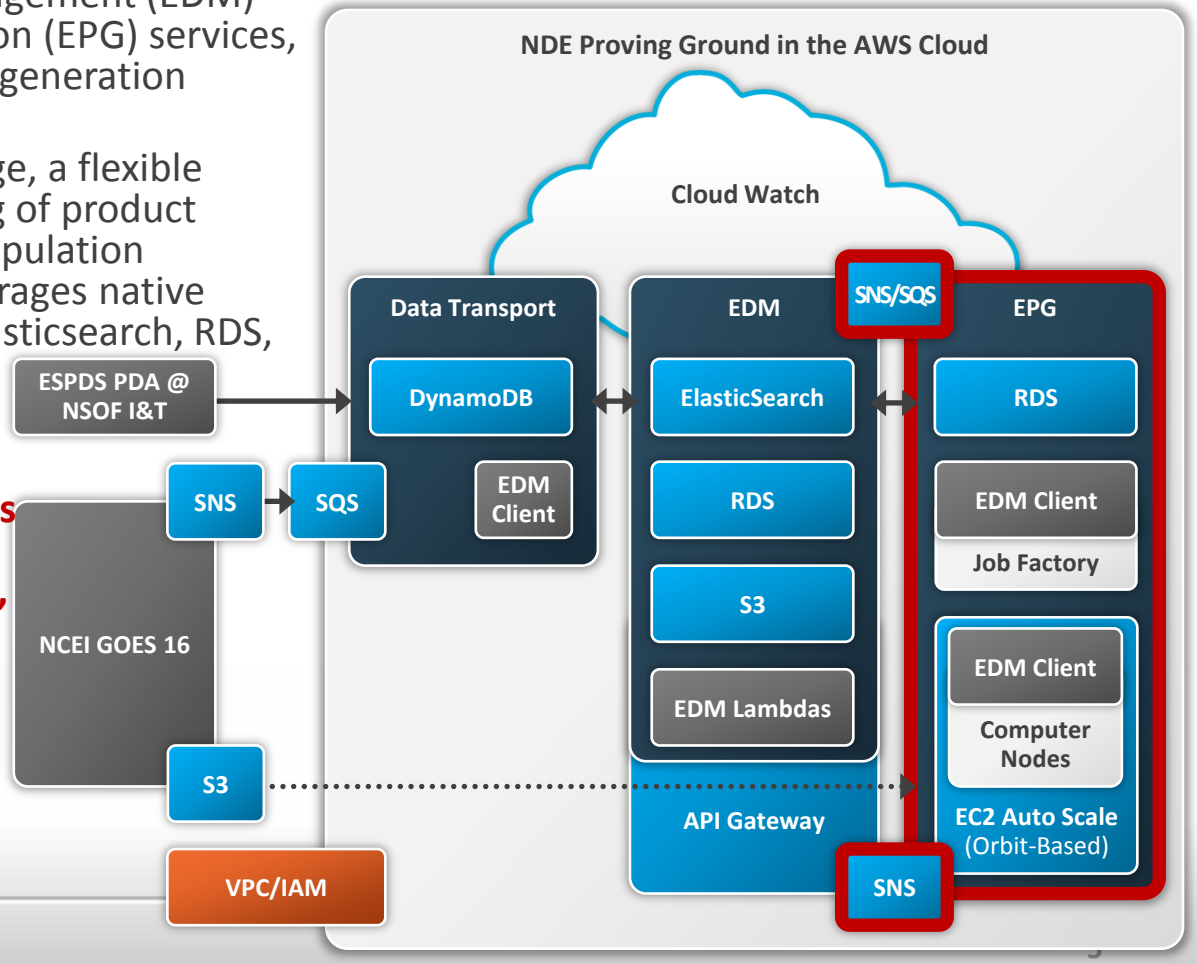
- As part of NOAA's Environmental Satellite Processing and Distribution System (ESPDS) program, Solers created a cloud platform for satellite data management and processing, known as the NOAA Data Exploitation (NDE) Proving Ground.
- Re-factoring of NOAA's on-premise operational NDE system to use AWS cloud services.
- Consists of Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services, with a subset of NOAA's product generation algorithms integrated.
- **EDM service provides data storage, a flexible and searchable inventory/catalog of product metadata, and science data manipulation through RESTful interfaces. Leverages native AWS cloud services including: Elasticsearch, RDS, S3, Lambda, and API Gateway.**
- EPG is capable of generating Level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.
- Data ingested:
  - GOES-16 data from the NOAA/NCEI Big Data Project (AWS S3 bucket).
  - S-NPP, JPSS-1, and GCOM-W data from ESPDS PDA at NSOF I&T.



SOLERS
A Peraton Company

NOAA Data Exploitation (NDE) Proving Ground Overview

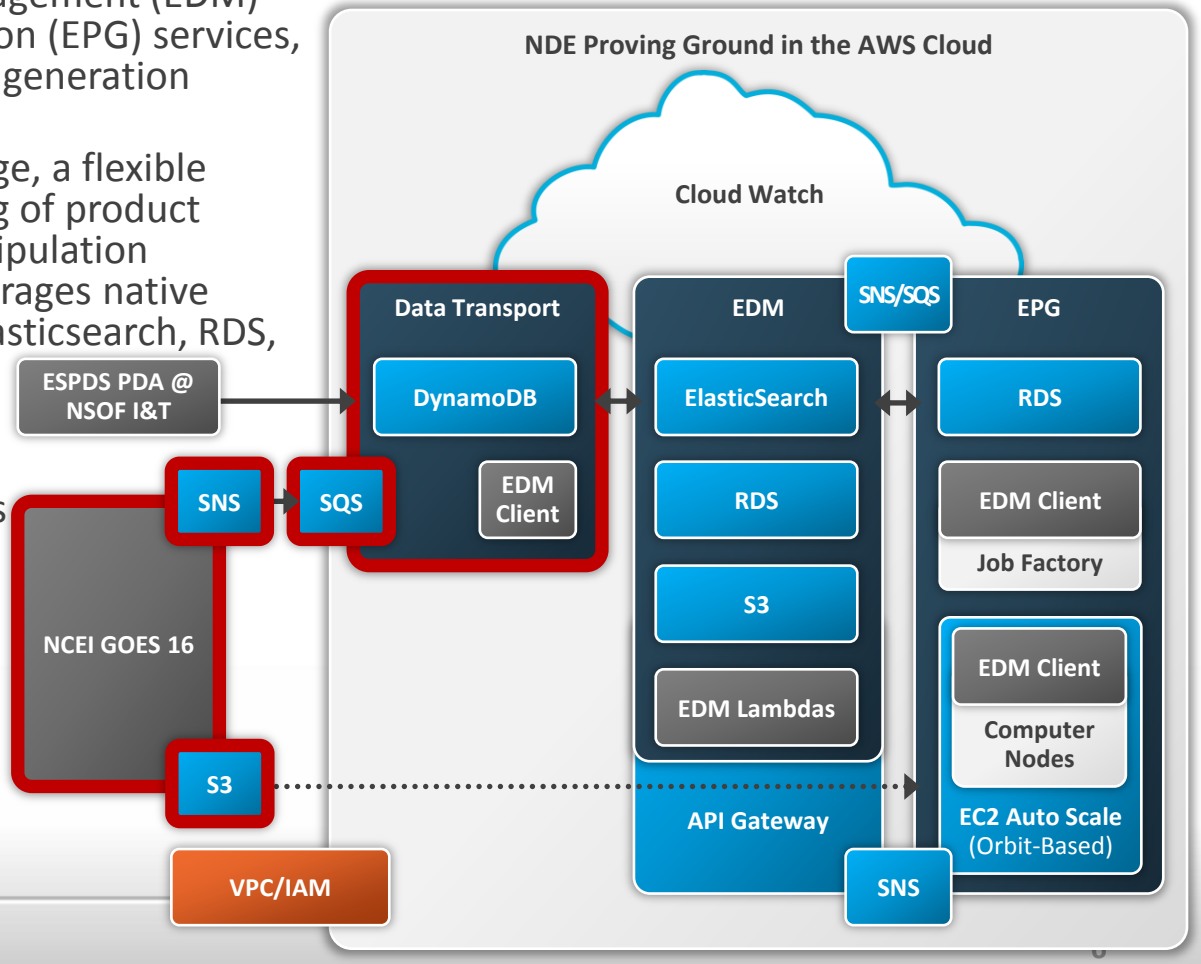
- As part of NOAA's Environmental Satellite Processing and Distribution System (ESPDS) program, Solers created a cloud platform for satellite data management and processing, known as the NOAA Data Exploitation (NDE) Proving Ground.
- Re-factoring of NOAA's on-premise operational NDE system to use AWS cloud services.
- Consists of Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services, with a subset of NOAA's product generation algorithms integrated.
- EDM service provides data storage, a flexible and searchable inventory/catalog of product metadata, and science data manipulation through RESTful interfaces. Leverages native AWS cloud services including: Elasticsearch, RDS, S3, Lambda, and API Gateway.
- **EPG is capable of generating Level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.**
- Data ingested:
  - GOES-16 data from the NOAA/NCEI Big Data Project (AWS S3 bucket).
  - S-NPP, JPSS-1, and GCOM-W data from ESPDS PDA at NSOF I&T.



SOLERS
A Peraton Company

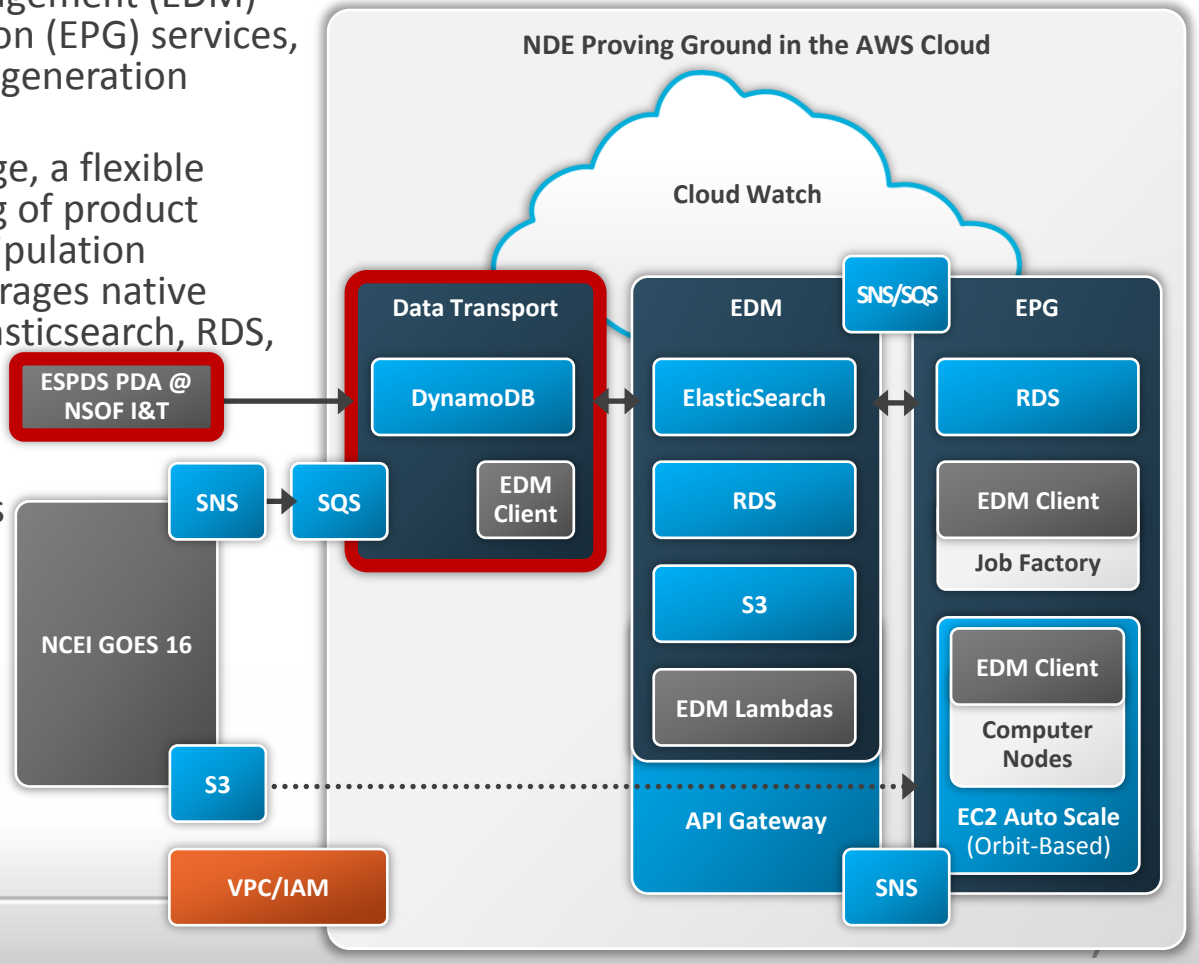
NOAA Data Exploitation (NDE) Proving Ground Overview

- As part of NOAA's Environmental Satellite Processing and Distribution System (ESPDS) program, Solers created a cloud platform for satellite data management and processing, known as the NOAA Data Exploitation (NDE) Proving Ground.
- Re-factoring of NOAA's on-premise operational NDE system to use AWS cloud services.
- Consists of Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services, with a subset of NOAA's product generation algorithms integrated.
- EDM service provides data storage, a flexible and searchable inventory/catalog of product metadata, and science data manipulation through RESTful interfaces. Leverages native AWS cloud services including: Elasticsearch, RDS, S3, Lambda, and API Gateway.
- EPG is capable of generating Level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.
- **Data ingested:**
  - **GOES-16 data from the NOAA/NCEI Big Data Project (AWS S3 bucket).**
  - S-NPP, JPSS-1, and GCOM-W data from ESPDS PDA at NSOF I&T.



NOAA Data Exploitation (NDE) Proving Ground Overview

- As part of NOAA's Environmental Satellite Processing and Distribution System (ESPDS) program, Solers created a cloud platform for satellite data management and processing, known as the NOAA Data Exploitation (NDE) Proving Ground.
- Re-factoring of NOAA's on-premise operational NDE system to use AWS cloud services.
- Consists of Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services, with a subset of NOAA's product generation algorithms integrated.
- EDM service provides data storage, a flexible and searchable inventory/catalog of product metadata, and science data manipulation through RESTful interfaces. Leverages native AWS cloud services including: Elasticsearch, RDS, S3, Lambda, and API Gateway.
- EPG is capable of generating Level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.
- **Data ingested:**
  - GOES-16 data from the NOAA/NCEI Big Data Project (AWS S3 bucket).
  - **S-NPP, JPSS-1, and GCOM-W data from ESPDS PDA at NSOF I&T.**



SOLERS
A Peraton Company

NOAA Data Exploitation (NDE) Proving Ground Objectives

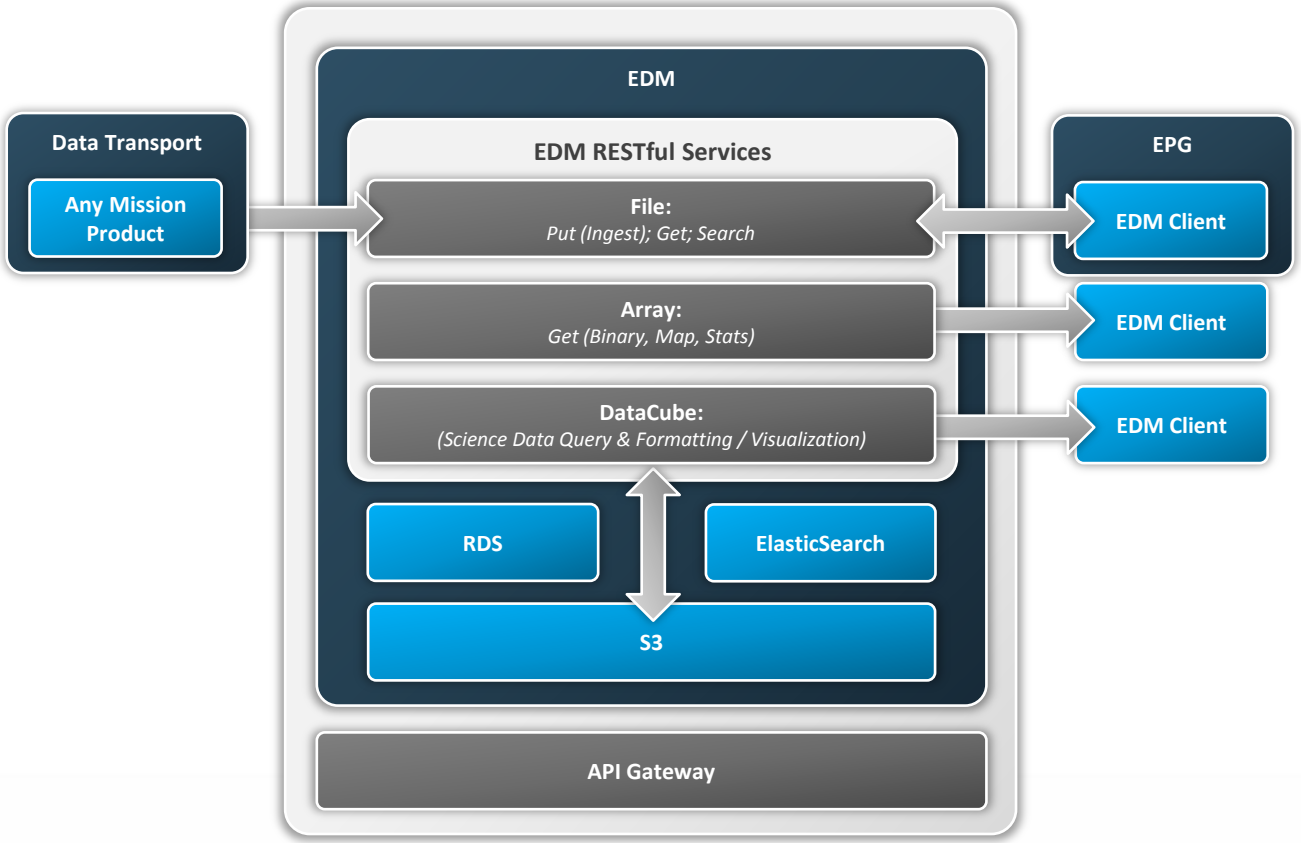
Primary Objectives:

- To leverage the flexibility and agility provided by a cloud environment to prototype candidate architectures and implementations for EDM and EPG services, and evaluate them for efficacy, performance, scalability, and maintainability.
- To demonstrate the flexibility of the proposed EPG service to execute multiple types of algorithms, such as existing ESPDS NDE 2.0 product algorithms, JPSS Risk Reduction algorithms, newer "NOAA Enterprise Algorithm" implementations of legacy products, and GOES-R L2+ product algorithms.
- To assess the cost of running these algorithms in a cloud environment.

Secondary Objectives:

- To consider how cloud-hosted EDM and EPG services could be used for collaboration and integration of future product generation algorithms, both within NOAA and with collaborative research organizations.
- To identify cost breakpoints for technology, ingress & egress, performance, etc.

EDM Overview



- RESTful Data Services
- Supports comprehensive access and manipulation of multi-mission science content
- Defines products across multiple missions
- Supports ingest, access, and analysis of products at multiple layers:
  - File
  - Array (i.e., access a specific array of a file only)
  - Data Cube (provides a Relational View and Query capability of science content that allows for filtering, sub-setting, down-sampling of aggregations across enterprise data holdings)
- Analysis Services are “attached” to the Data Services, examples:
  - Imaging
  - Mapping
  - Statistical Analysis/Summary

EDM Metadata Services

Why a Rich Metadata Environment?

- Defines a common data abstraction that becomes a foundation for development of Data Services independent of Mission/Product implementation
- Provides enhanced discovery capabilities
  - Full text and spatial search of total metadata content
- Provides a scaffolding for Enhanced Data Services
- Provides quality control
  - Array level summary statistics of science content could be stored in the JSON document for comparison against seasonal/regional statistics providing automated identification of science content deviating from an expected baseline

EDM Metadata Enhancements

JPSS Example JSON document:

GOES-16 Example JSON document:

Consistent Across Enterprise

"edmCore" : {

```
        "platformNames" : "NPP",
        "productShortName" : "CrIS-FS-SDR",
        "fileId" : 33042832,
        "fileName" :
"SCRIF_npp_d20180918_t2105439_e2106137_b35717_c20180918224610354086_niic_int.h5
",
        "fileStartTime" : "20180918T210543.900Z",
        "fileEndTime" : "20180918T210613.700Z",
        "fileInsertTime" : "20180920T210029.403Z",
        "fileSpatialArea" : { ... }
      },
```

"edmCore" : {

```
  "fileId" : 33194512,
  "fileName" : "OR_ABI-L1b-RadM2-M3C02_G16_s20182601757511_e20182601757568_c20182601758001.nc",
  "productShortName" : "ABI-L1b-RadM2-C02",
  "fileSpatialArea" : { ... },
  "fileStartTime" : "20180917T175751.100Z",
  "fileEndTime" : "20180917T175756.800Z",
  "fileInsertTime" : "20180920T235401.526Z",
  "platformNames" : ["G16" ]
 },
```

Unique to Product

"objectMetadata" : {

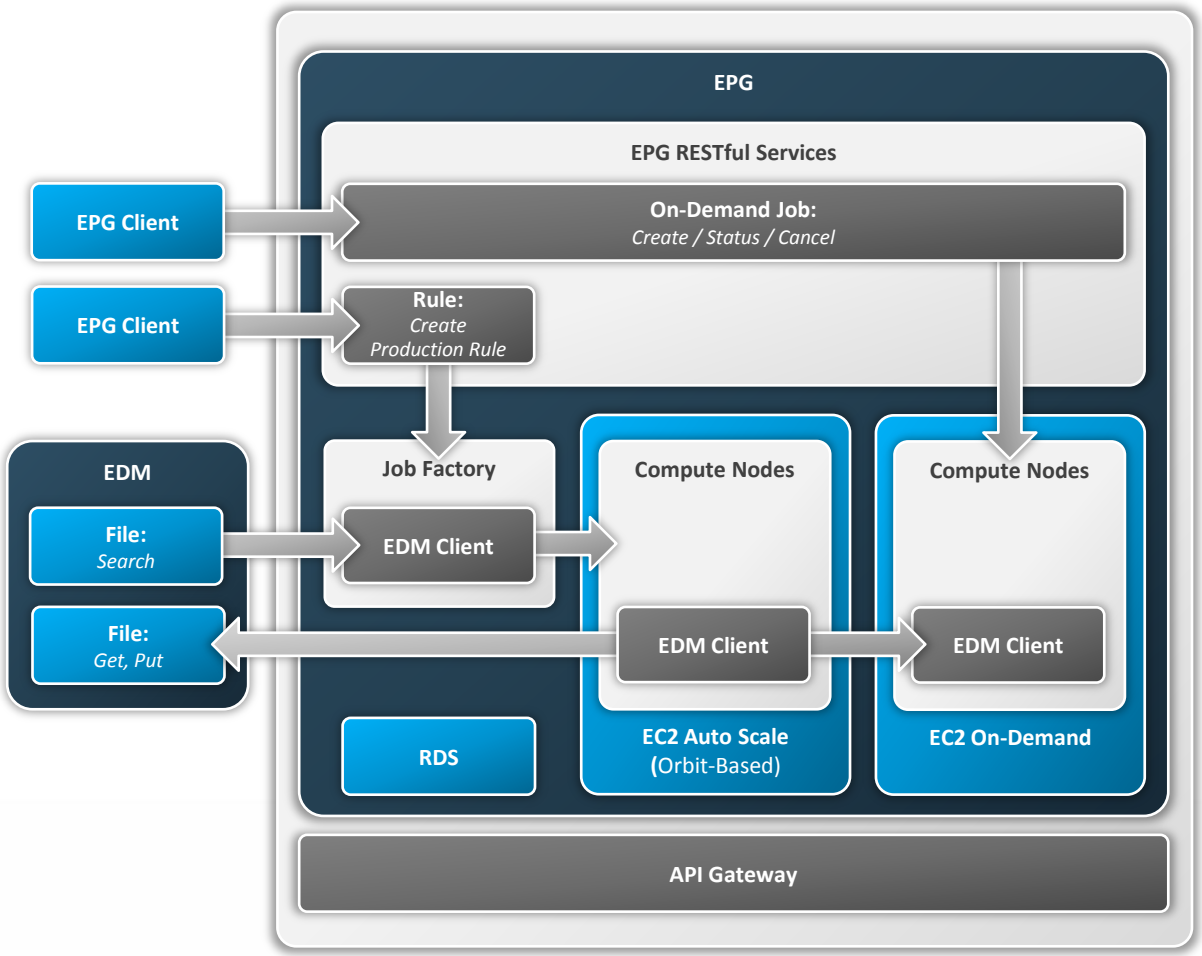
```
      "attributes" : {
        "Distributor" : "nii-",
        "Mission_Name" : "S-NPP/JPSS",
        "N_Dataset_Source" : "nii-",
        "N_GEO_Ref" :
"GCRSO_npp_d20180918_t2105439_e2106137_b35717_c20180918224610385032_niic_int.
h5",
        "N_HDF_Creation_Date" : "20180918",
        "N_HDF_Creation_Time" : "224610.354086Z",
        "Platform_Short_Name" : "NPP"
      },
      "datasets" : { },
      "datatypes" : { },
      "All_Data" : {
       "CrIS-FS-SDR_All" : {
         "datasets" : {
          "DS_SpectralStability" : {
           "datatype" : "float64",
           "group" : "/All_Data/CrIS-FS-SDR_All",
           "size" : 216,
           "shape" : [4, 2, 9, 3]
          },
```

"objectMetadata" : {

```
  "attributes" : {
   "naming_authority" : "gov.nesdis.noaa",
   "Conventions" : "CF-1.7",
   "Metadata_Conventions" : "Unidata Dataset Discovery v1.0",
   "standard_name_vocabulary" : "CF Standard Name Table (v25, 05 July 2013)",
   "institution" : "DOC/NOAA/NESDIS > U.S. Department of Commerce...",
   "project" : "GOES",
   "production_site" : "RBU",
   "production_environment" : "OE",
   "spatial_resolution" : "0.5km at nadir",
   "orbital_slot" : "GOES-East",
   "platform_ID" : "G16",
   "instrument_type" : "GOES R Series Advanced Baseline Imager",
...
  },
  "dimensions" : {
   "y" : 2000,
   "x" : 2000,
   "number_of_time_bounds" : 2,
   "band" : 1,
   "number_of_image_bounds" : 2,
   "num_star_looks" : 24
  },
  "variables" : {
   "Rad" : {
    "datatype" : "int16",
    "shape" : [ 2000, 2000],
    "size" : 4000000,
    "dimensions" : ["y", "x" ],
    "attributes" : {
     "_FillValue" : 4095,
     "long_name" : "ABI L1b Radiances",
     "standard_name" : "toa_outgoing_radiance_per_unit_wavelength",
...
    }
   },
   "DQF" : {
```

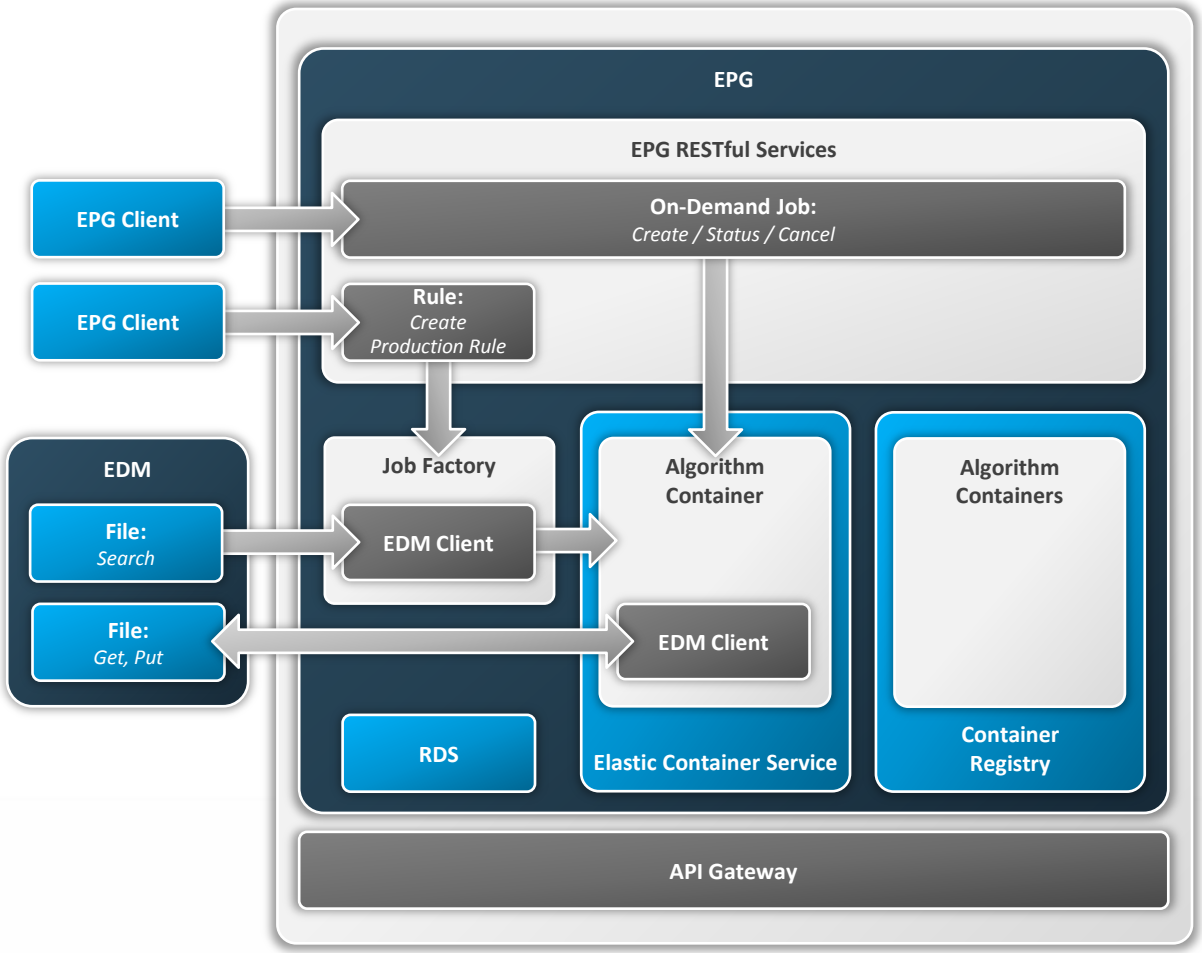
EDM stores one JSON metadata document per file. Each document contains an edmCore section and an objectMetadata section.

EPG Overview



- RESTful Product Generation Services
- Baseline NDE PG Capabilities:
  - Algorithm and Production Rule Definition
  - Event Driven Job Creation and Load Management
  - Algorithm Execution (runs any algorithm as an executable as specified in the Production Rule)
- Enhanced PG Services:
  - Access to EDM RESTful API
    - Common Data Access Interfaces
    - Enhanced Data Availability / Selection
  - Data Availability Subscription/Notification
  - On Demand Production Rule Creation
  - On-Demand Job Creation

EPG Support for Containerized Algorithms



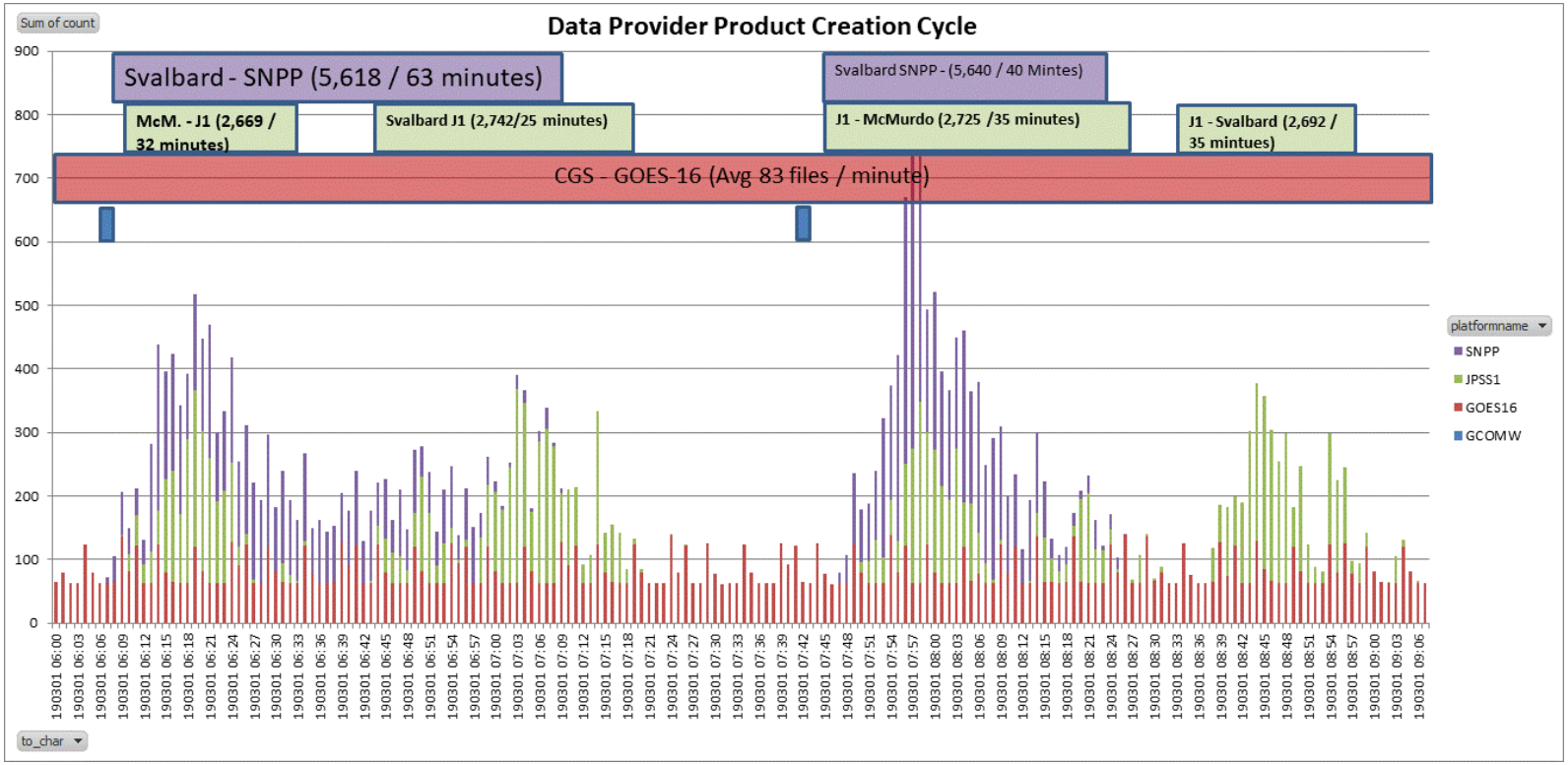
- Support for Algorithm Containers
  - Received and integrated "containerized" versions of 2 algorithms from STAR (NOAA's Center for Satellite Applications and Research)
    - Algorithm libraries/files and executable are packaged up and invoked as an OCI-compliant container image
  - Added an Algorithm Container Registry and the AWS Elastic Container Service (ECS) to support this integration
  - Containerized algorithms produced comparable products to the standard executable versions, but runtimes were slightly longer

Standard Executable vs. Containerized Algorithm Runtimes

Unit Name	Executable Mean Runtime (seconds)	Container Mean Runtime (seconds)	Difference %
VPW PRODUCT S-NPP	355.31	358.02	0.76%
VPW PRODUCT JPSS1	241.03	243.99	1.23%
VPW REMAP S-NPP	131.92	146.09	10.74%
VPW REMAP JPSS1	128.56	139.61	8.59%
Cloud Mask S-NPP	256.73	270.56	5.39%
Cloud Mask JPSS1	258.93	267.58	3.34%

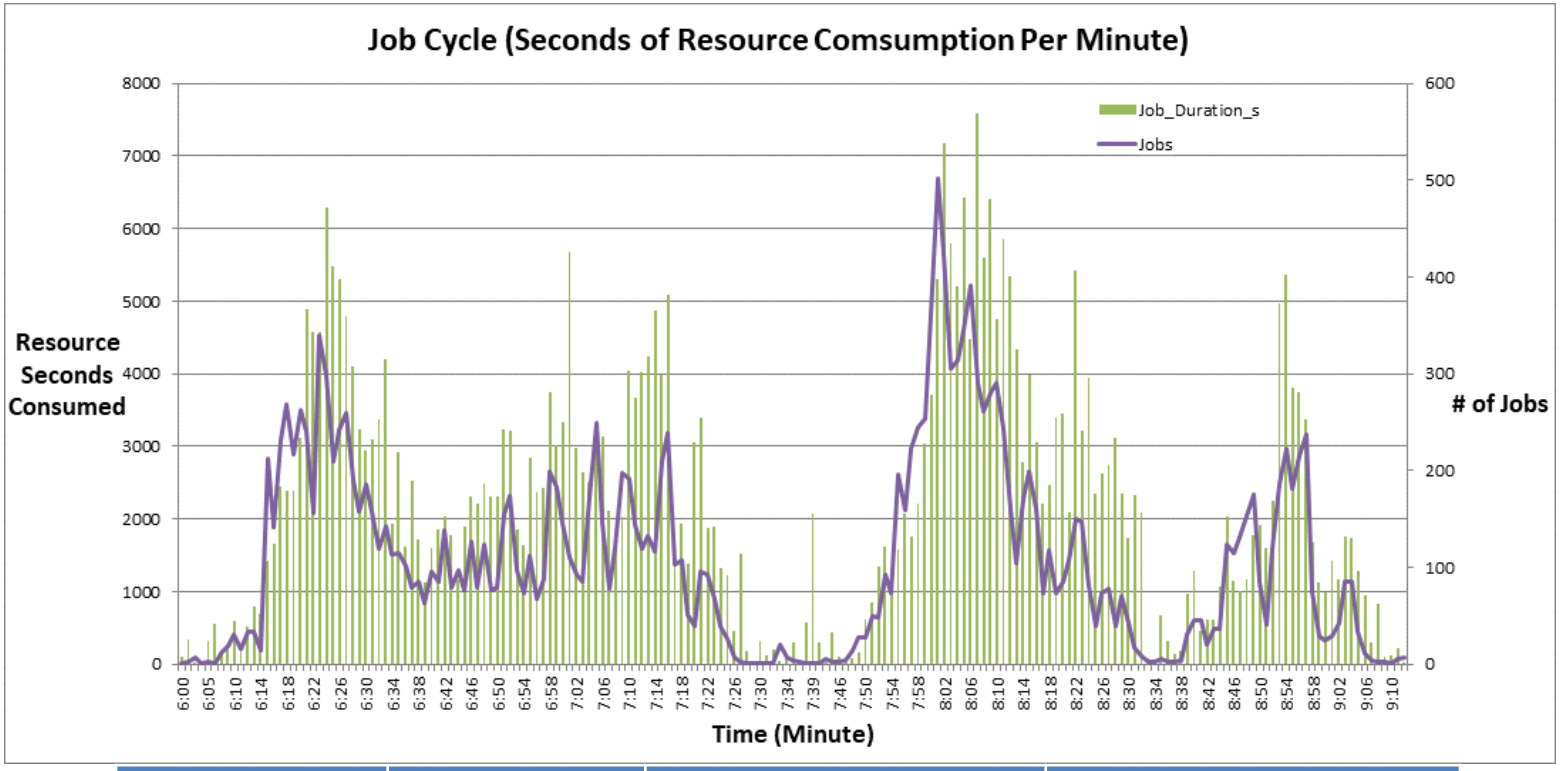
- Standard Executable vs. Containerized algorithm runtimes for VIIRS Polar Winds (VPW) and Cloud Mask algorithms applied to both the S-NPP and JPSS-1 ingested data
- VPW includes both generating the product, and performing a remapping of the generated product

21-Day Test: Daily Data Ingest Cycle and Volumes



Satellite	Ingested Data Types	Ingested Data Volume (Files per Day)	Ingested Data Volume (GB per Day)
GCOM-W	1	14	1.71
GOES-16	64	52,224	132.89
JPSS-1	65	77,756	886.18
S-NPP	73	84,256	888.52
TOTAL	**203**	**214,250**	**1,909.30**

21-Day Test: Daily Product Generation Cycle and Volumes



Satellite	Generated Product Types	Generated Product Volume (Files per Day)	Generated Product Volume (GB per Day)
GCOM-W	32	457	31.14
GOES-16	240	181,877	918.26
JPSS-1	250	137,552	1,920.12
S-NPP	303	164,103	2,017.93
TOTAL	**825**	**483,989**	**4,887.45**

Estimated Yearly AWS Compute and Storage Costs

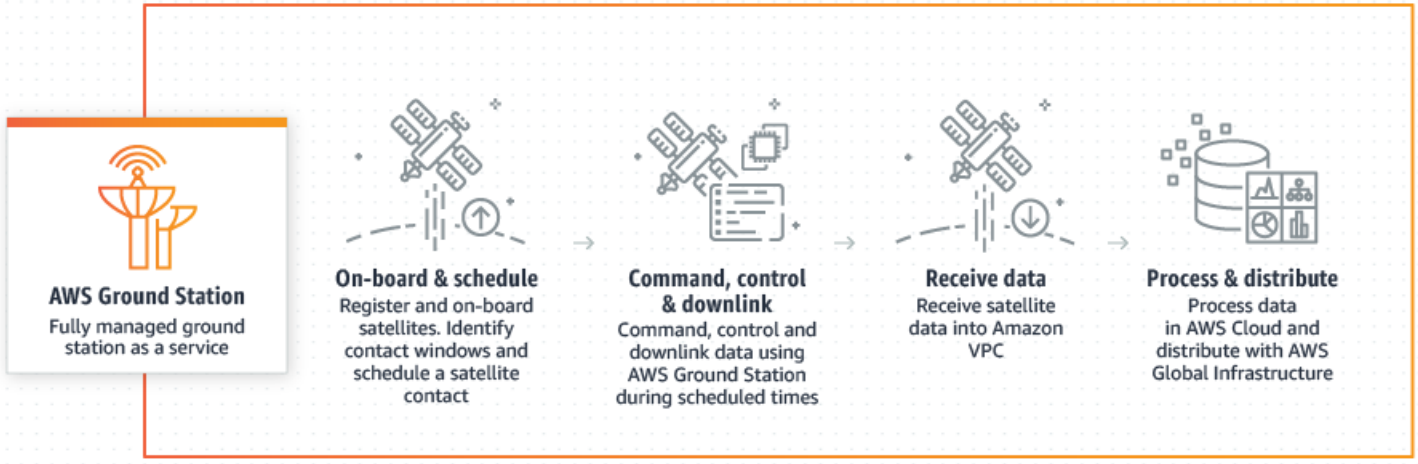
Satellite	Estimated Yearly Compute and Storage Costs
GCOM-W	$2,171.37
GOES-16	$79,690.10
JPSS-1	$59.483.85
S-NPP	$59,483.85
TOTAL	**$200,829.17**

For comparison, NOAA's On-Premise Operational NDE System costs over $1M per year in hardware maintenance and COTS software licensing/support, with additional costs incurred during tech refresh cycles prior to vendor-announced End of Life (EOL) / End of Support (EOS) dates.

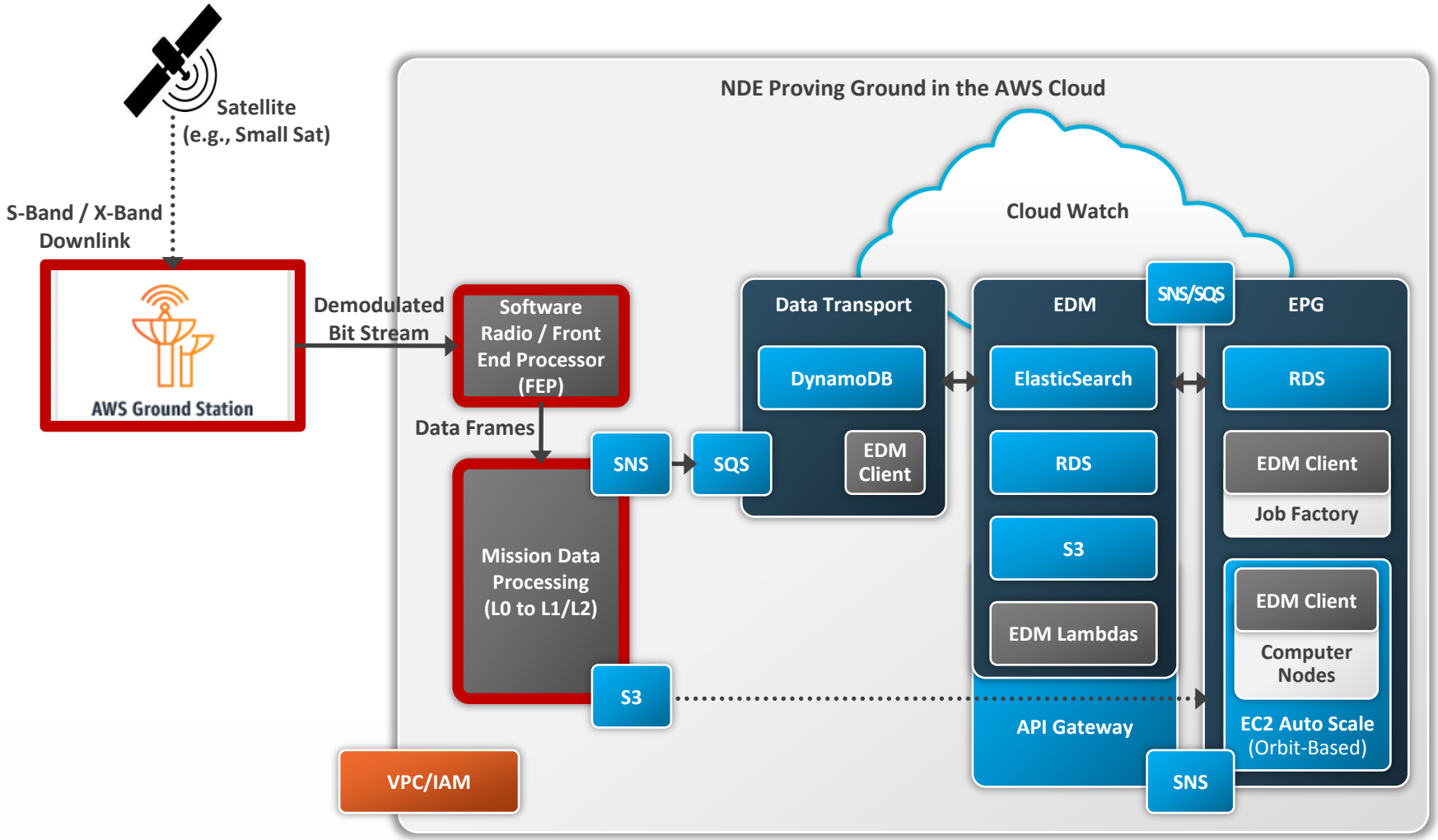
AWS Ground Station

https://aws.amazon.com/ground-station

- Enables satellite operators to control and ingest data from orbiting satellites without having to buy or build satellite ground station infrastructure
- Integrates the ground station equipment like antennas, digitizers, and modems into the AWS regions around the world
  - General availability in US-West-2 (Oregon) and US-East-2 (Ohio), with 12 planned Ground Station locations in 2019 *(per AWS Public Sector Summit in Washington, DC, 10-12 June 2019)*
- Satellite operators can simply onboard their satellites and schedule time to communicate with them using AWS Ground Station
- Satellite operators have the option of conducting all of their satellite operations within the AWS cloud, including the storing and processing of the satellite data and delivering products using AWS services, or use AWS Ground Station just to downlink the satellite data and transport it to their own processing center



AWS Ground Station and NDE Proving Ground Notional Integrated Architecture



Conclusion

- NDE Proving Ground architecture provides a scalable AWS cloud platform for satellite data processing that has been shown to be capable of realizing significant cost annual cost savings over operational on-premise equivalents (i.e., NOAA use cases)
- When paired with AWS Ground Station, it can provide the capabilities necessary to perform scheduled downlink and mission data processing for Small Satellites, without requiring any on-premise hardware or infrastructure
- Representing data processing and product generation algorithms as “micro-services” is recommended especially for “new” small satellite data/product use cases
  - Keeping the algorithm package as small and lightweight as possible will ease its portability and configuration management, and maximize the efficiency of managing its execution within the EPG services
  - Provided that the processing logic fits within the AWS-enforced restrictions, representing data processing and product generation algorithms as AWS Lambda functions would be the most efficient and cost-effective mechanism of integrating algorithms into the EPG services
  - In cases where this is not possible, then implementing and integrating the algorithms as containers would be the next recommended method to preserve the “micro-services” concept, maximize the algorithm’s portability, and ease the algorithm’s configuration management

Questions