SOLERS
An Employee Owned Company

NOAA/NESDIS Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) Proving Ground in the AWS Cloud

GSAW 2019
Cloud Computing and Big Data Technologies for Ground Systems WG

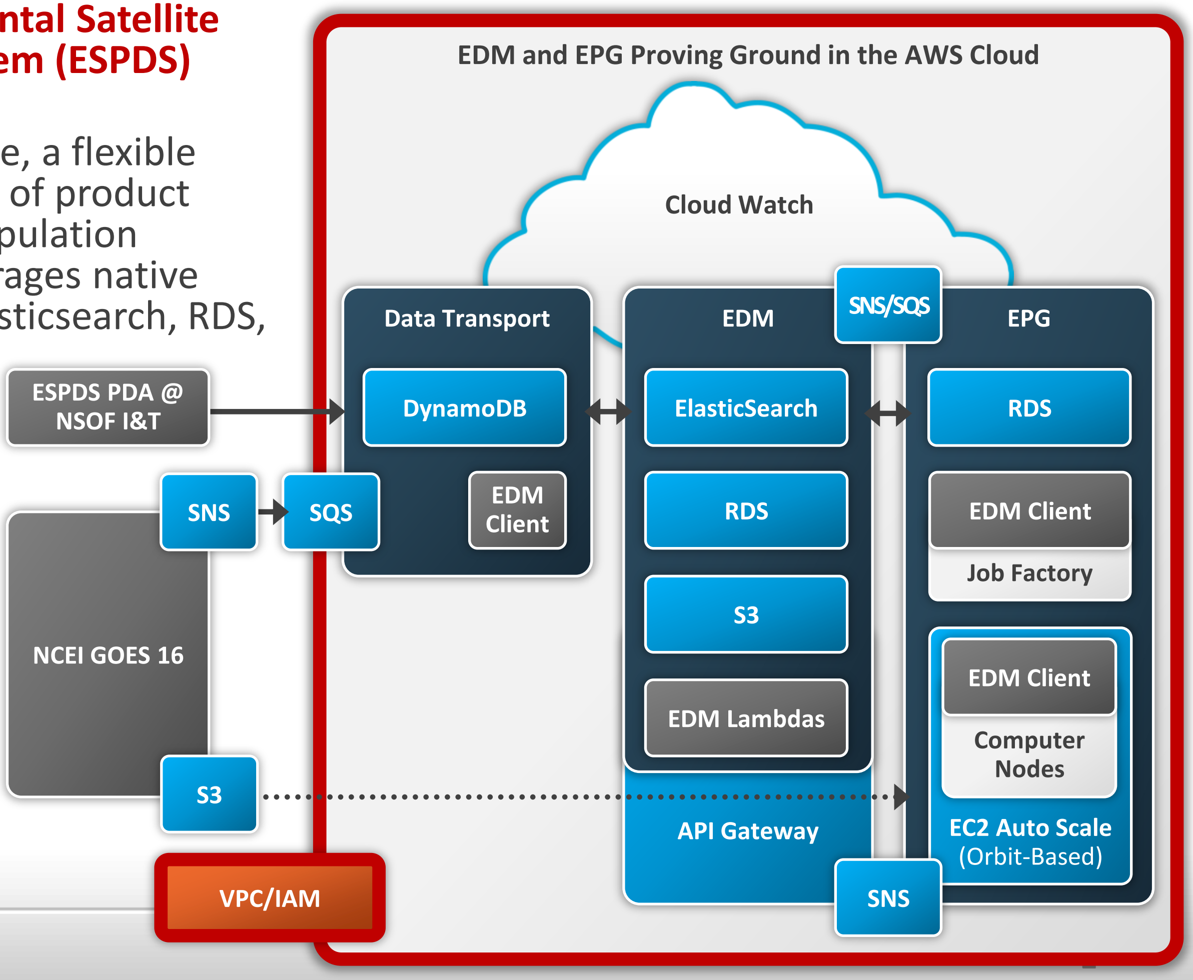
Rich Baker
Chief Architect
Solers, Inc.
Email: richard.baker@solers.com
Phone: 240-790-3338

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

www.solers.com

EDM and EPG Proving Ground Overview

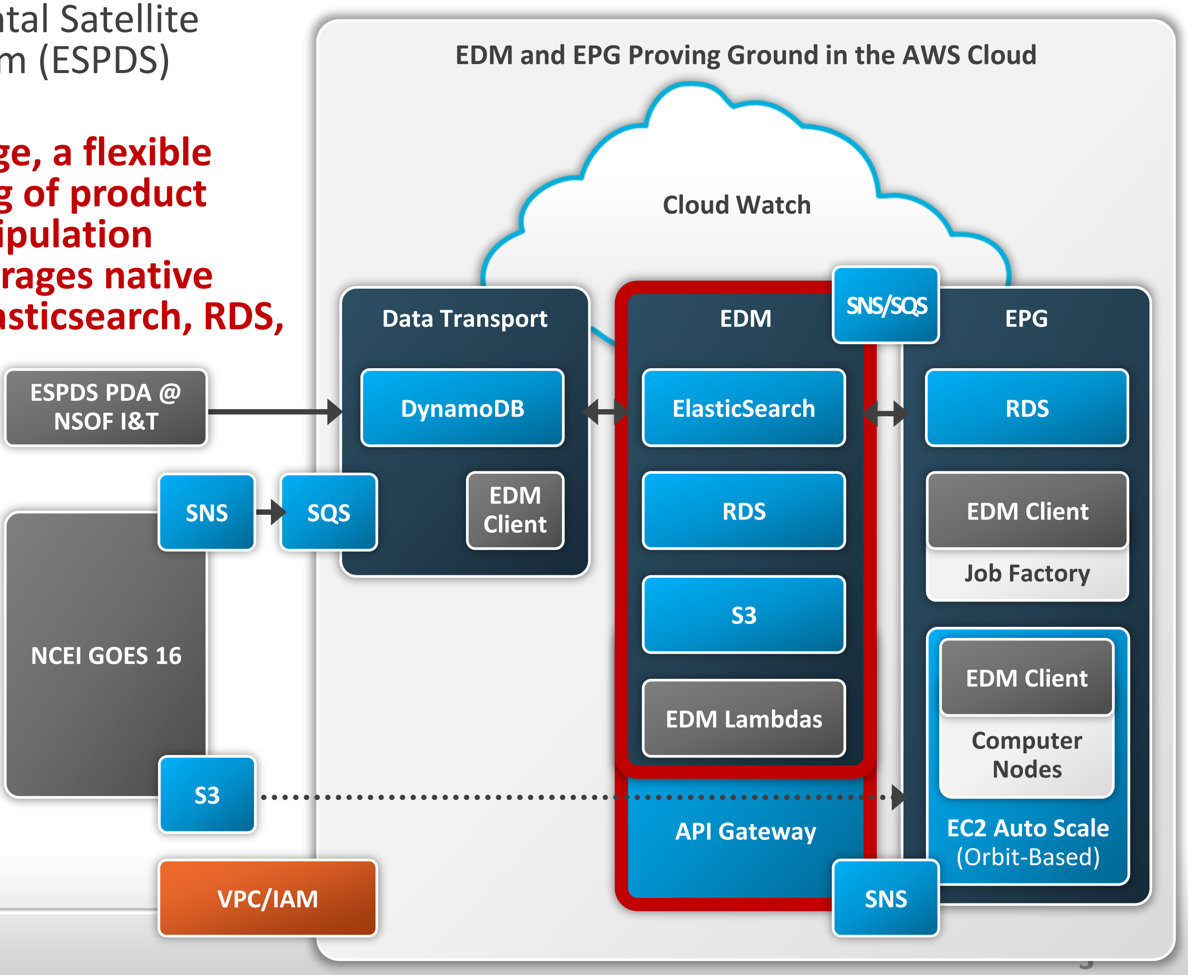
- **Solers created a Proving Ground for Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services in a FedRAMP-approved Amazon Web Services (AWS) cloud environment, leveraging native AWS cloud services and NESDIS product generation algorithms.**
- **Developed under the Environmental Satellite Processing and Distribution System (ESPDS) contract.**
- 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 NESDIS level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.
- Data currently being 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
An Employee Owned Company

EDM and EPG Proving Ground Overview

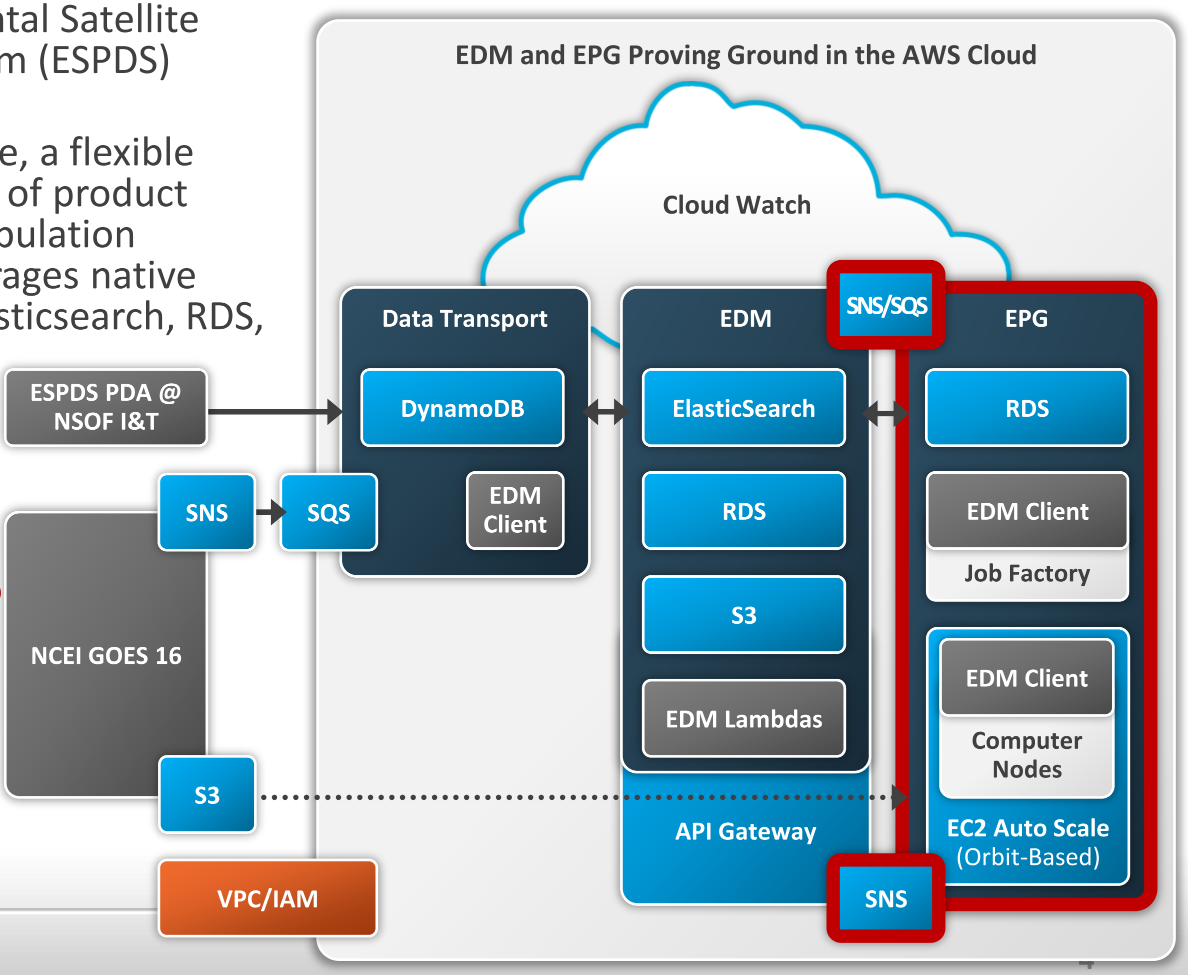
- Solers created a Proving Ground for Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services in a FedRAMP-approved Amazon Web Services (AWS) cloud environment, leveraging native AWS cloud services and NESDIS product generation algorithms.
- Developed under the Environmental Satellite Processing and Distribution System (ESPDS) contract.
- **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 NESDIS level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.
- Data currently being 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
An Employee Owned Company

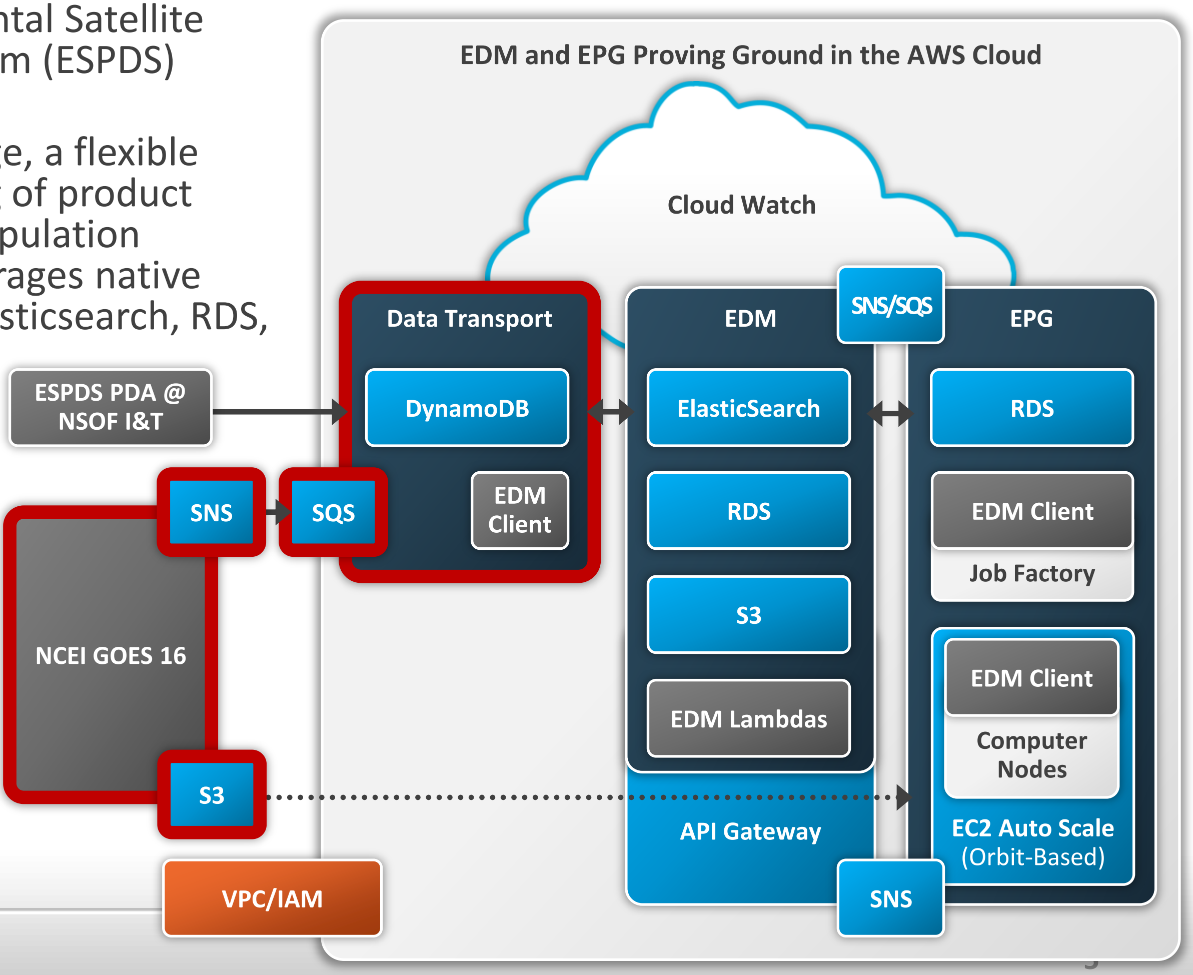
EDM and EPG Proving Ground Overview

- Solers created a Proving Ground for Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services in a FedRAMP-approved Amazon Web Services (AWS) cloud environment, leveraging native AWS cloud services and NESDIS product generation algorithms.
- Developed under the Environmental Satellite Processing and Distribution System (ESPDS) contract.
- 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 NESDIS level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto Scaling, RDS, SNS, and SQS.**
- Data currently being 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.



EDM and EPG Proving Ground Overview

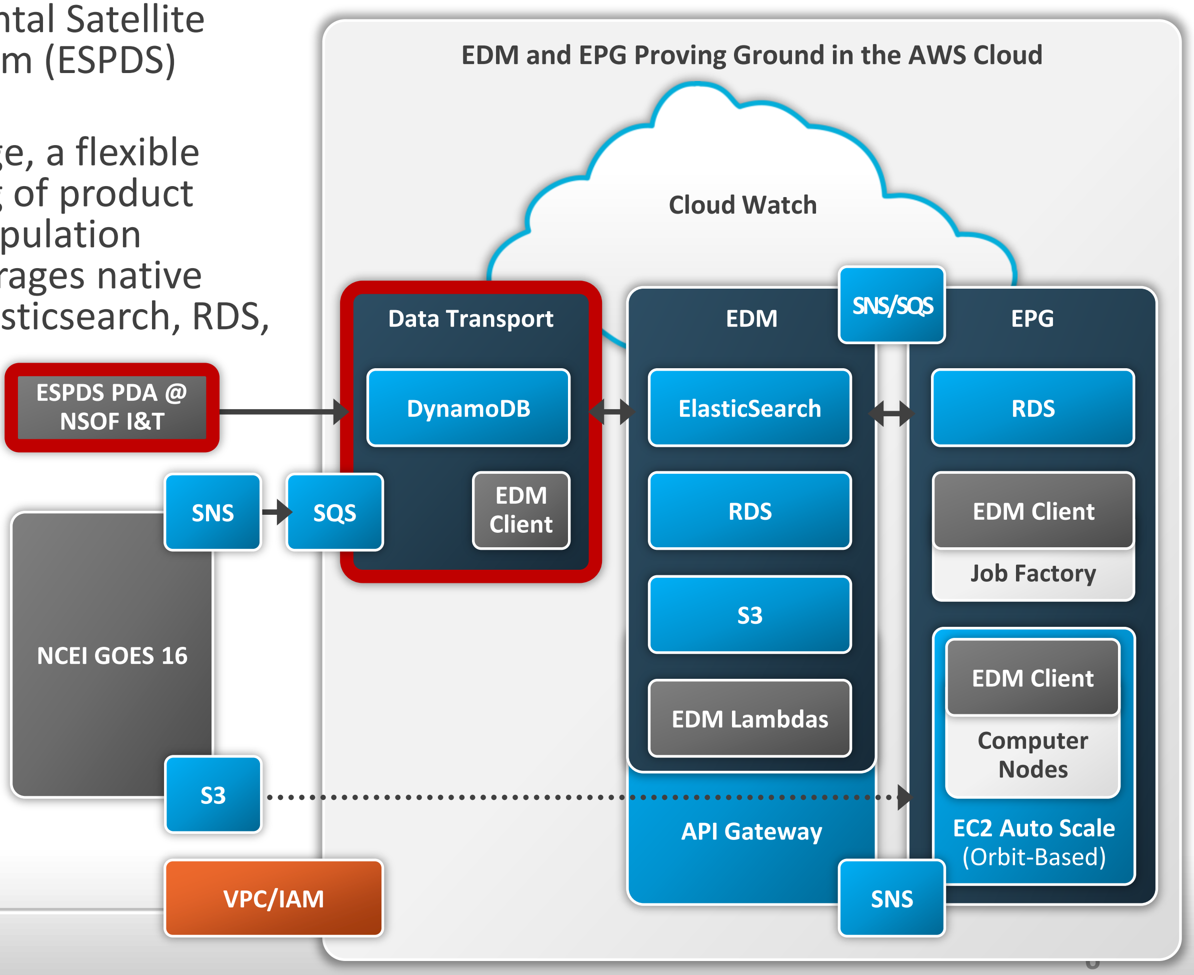
- Solers created a Proving Ground for Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services in a FedRAMP-approved Amazon Web Services (AWS) cloud environment, leveraging native AWS cloud services and NESDIS product generation algorithms.
- Developed under the Environmental Satellite Processing and Distribution System (ESPDS) contract.
- 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 NESDIS level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.
- **Data currently being 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 An Employee Owned Company

EDM and EPG Proving Ground Overview

- Solers created a Proving Ground for Enterprise Data Management (EDM) and Enterprise Product Generation (EPG) services in a FedRAMP-approved Amazon Web Services (AWS) cloud environment, leveraging native AWS cloud services and NESDIS product generation algorithms.
- Developed under the Environmental Satellite Processing and Distribution System (ESPDS) contract.
- 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 NESDIS level 1+ sensor, science, and tailored product types. Leverages native AWS cloud services including: EC2 with Auto-Scaling, RDS, SNS, and SQS.
- **Data currently being 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
An Employee Owned Company

EDM and EPG 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, NESDIS/STAR Enterprise Algorithm implementations of legacy products, and GOES-R L2+ product algorithms.
- To assess the cost of running these algorithms in a cloud environment.

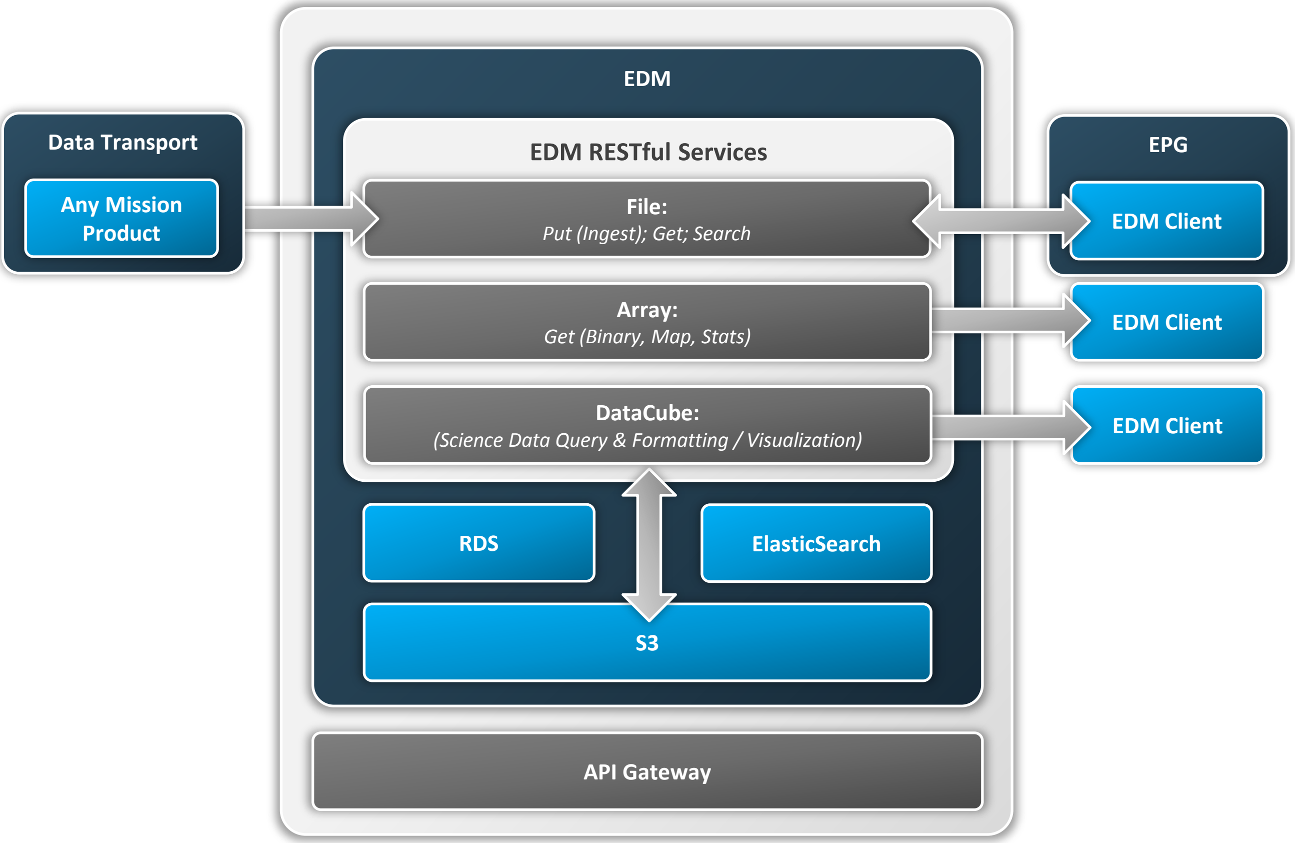
Secondary Objectives:

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

EDM and EPG Proving Ground Current Status

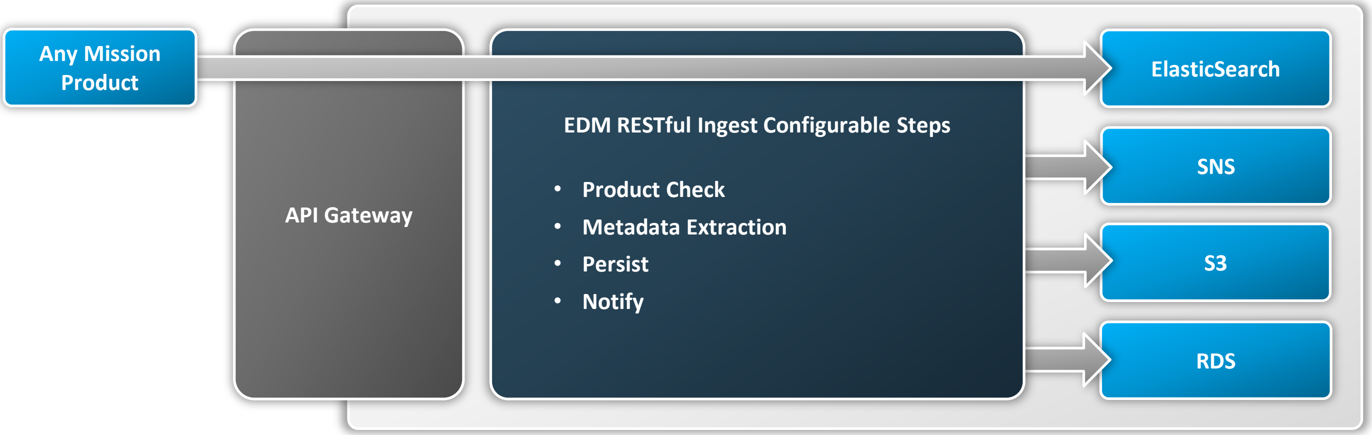
- EDM and EPG environments are established in the in the NOAA OCIO FedRAMP-approved AWS Cloud environment
  - Utilizes AWS Cloud services and existing science algorithms
  - Data feeds from ESPDS PDA at NSOF I&T (GCOM-W, JPSS-1, S-NPP) and NOAA/NCEI Big Data Project S3 Bucket (GOES-16)
- Products are being generated from Polar and Geo Missions, including:
  - **GCOM-W:** AMSR2-L1, GAASP
  - **JPSS-1:** Active Fire, JRR(Alpha), NUCAPs, OMPS, Tailoring, True-Color
  - **S-NPP:** ACSPO, Active Fire, GVF, JRR, MiRS, NUCAPS, OMPS, OMPS V8 TOS, Tropical Cyclone, SR, VH, VI, Polar Winds, Tailoring, True-Color
  - **GOES-16:** GOES-R L2 Products (~ half) via U-Wisconsin CSPP Package, DMW Algorithm (STAR), DMW BUFR, Tailoring
- In the process of coordinating with OSPO/STAR for cursory product quality analysis

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 Enhancements



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:

```
"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" : { … }
      },
```

```
"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]
            },
```

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

GOES-16 Example JSON document:

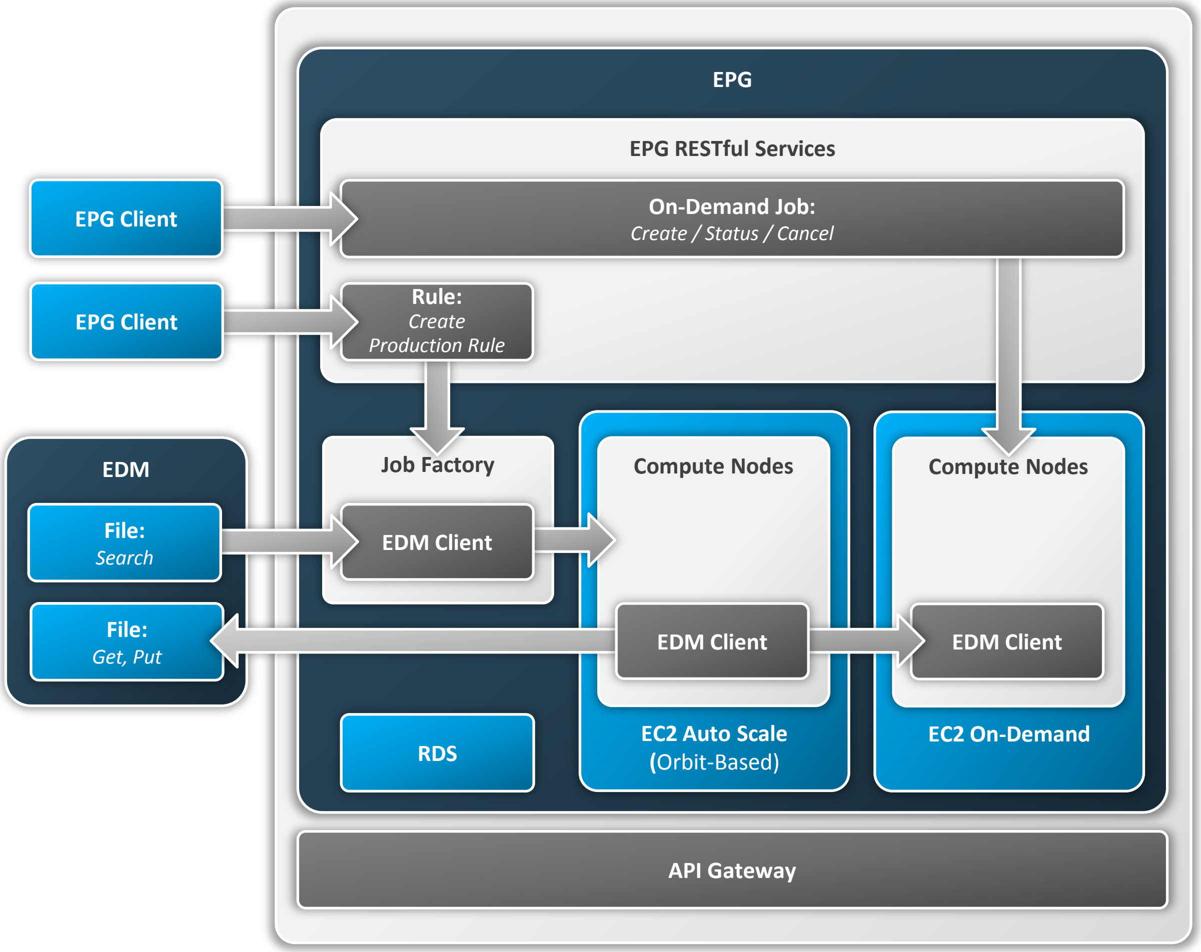
```
"edmCore" : {
      "fileId" : 33194512,
      "fileName" : "OR_ABI-L1b-RadM2-M3C02_G16_s20182601757511_e20182601757568_c20182601758001.nc",
      "productShortName" : "ABI-L1b-RadM2-C02",
      "fileSpatialArea" : { … },
      "fileStartTime" : "20180917T175751.100Z",
      "fileEndTime" : "20180917T175756.800Z",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

Consistent Across Enterprise

```
"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" : {
```

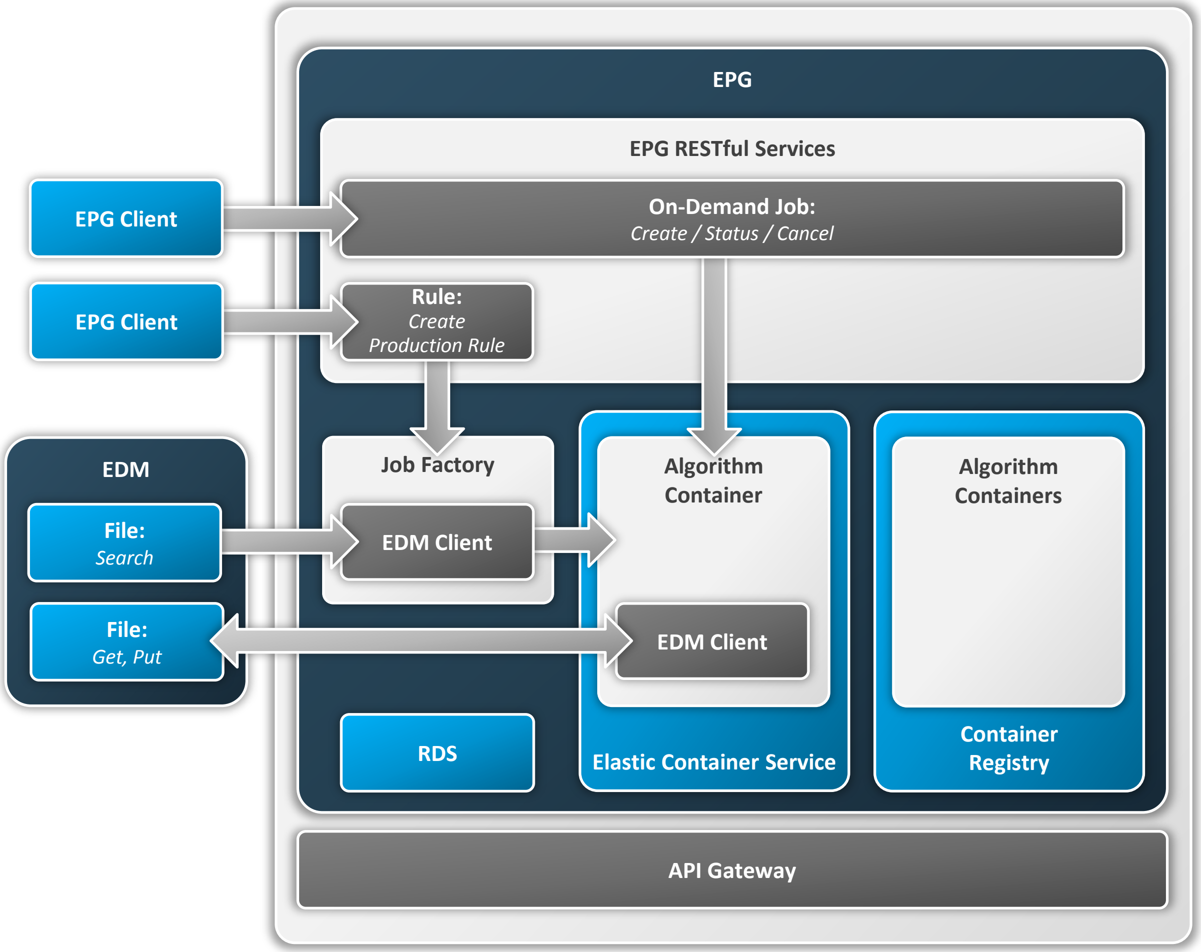
Unique to Product

EPG Overview



- RESTful Product Generation Services
- Current NDE PG Capabilities:
  - Algorithm and Production Rule Definition
  - Event Driven Job Creation and Load Management
- 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

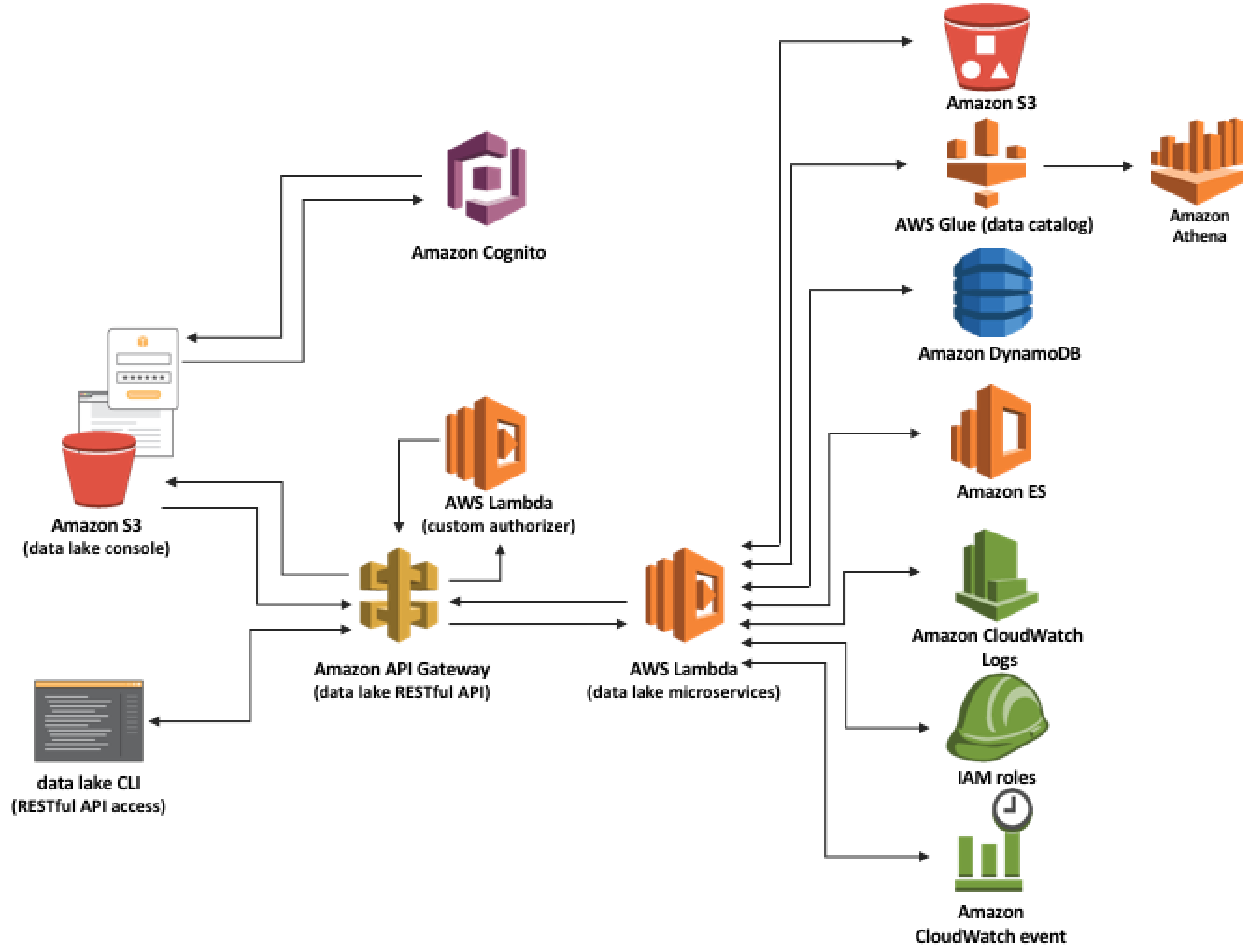
EPG Planned Modifications



- Support for Algorithm Containers
  - Will be receiving containerized versions of algorithms from STAR
  - Will add Algorithm Container Registry and Elastic Container Service (ECS) to the existing EPG capabilities
  - Evaluation of capabilities and limitations of ECS for containerized algorithms
  - Perform cost / performance comparison of container versus compute instance approach to EPG

Following the AWS Data Lake Architecture for Managing Big Data

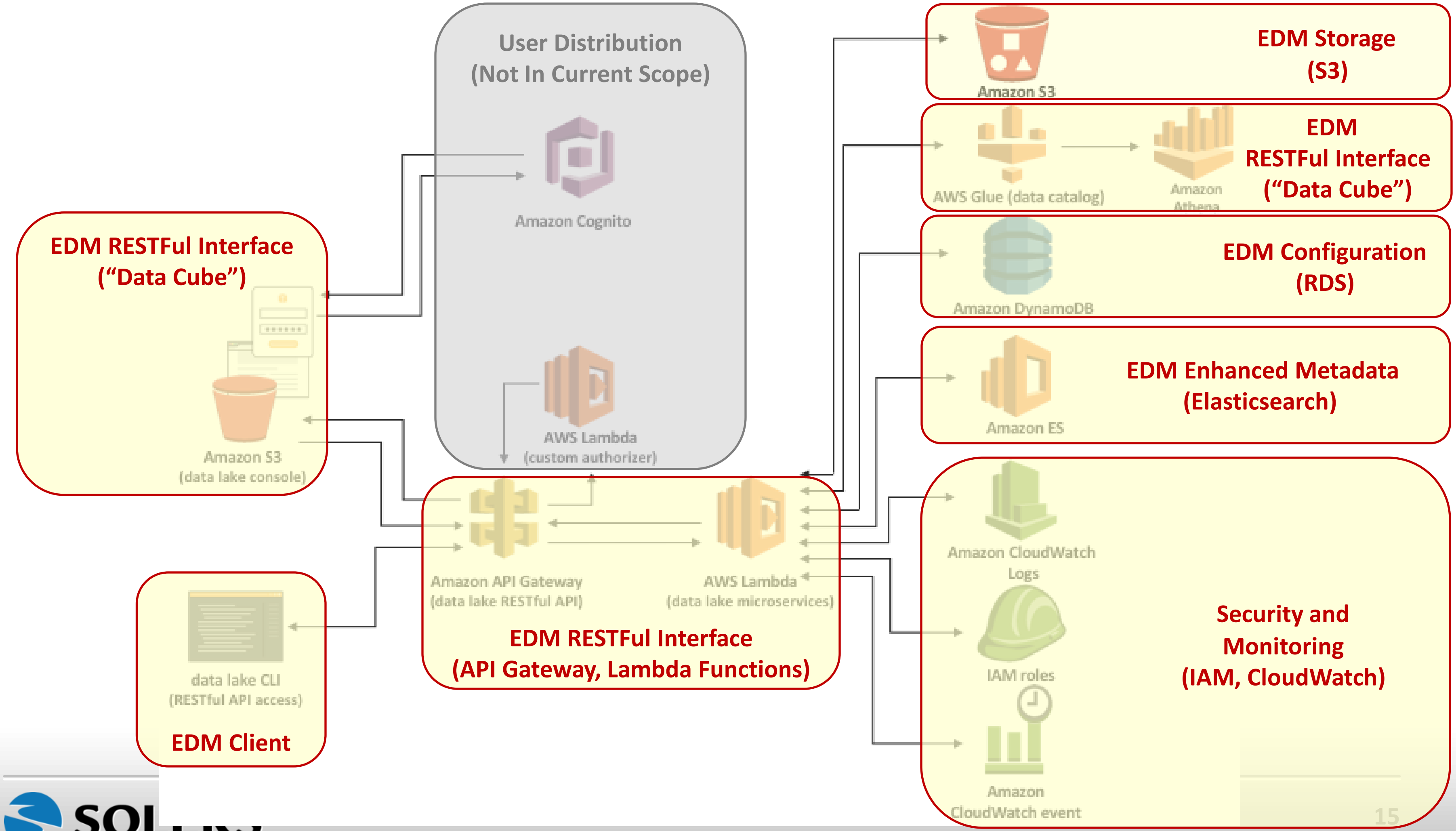
- ➢ Following a similar architecture as that of the AWS Data Lake solution published by Amazon (https://aws.amazon.com/answers/big-data/data-lake-solution)



SOLERS
An Employee Owned Company

Following the AWS Data Lake Architecture for Managing Big Data

- Following a similar architecture as that of the AWS Data Lake solution published by Amazon (https://aws.amazon.com/answers/big-data/data-lake-solution)



User Distribution
(Not In Current Scope)

EDM RESTFul Interface
(“Data Cube”)

EDM Client

EDM RESTFul Interface
(API Gateway, Lambda Functions)

EDM Storage
(S3)

EDM
RESTFul Interface
(“Data Cube”)

EDM Configuration
(RDS)

EDM Enhanced Metadata
(Elasticsearch)

Security and
Monitoring
(IAM, CloudWatch)

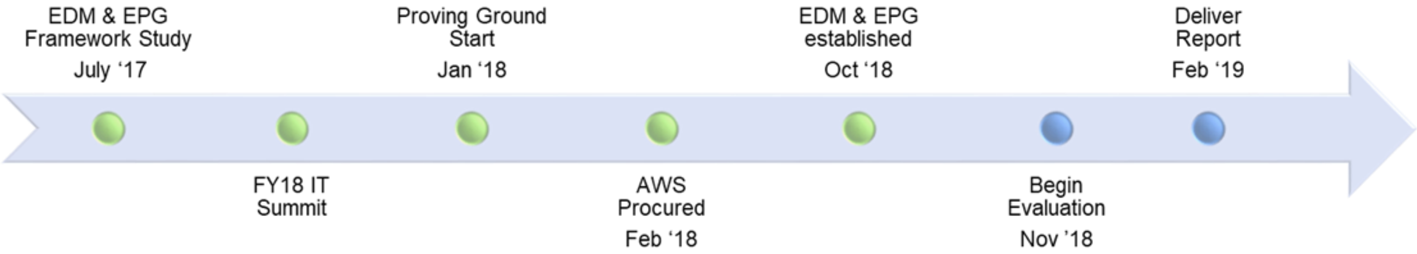
Data Volumes (GB/Day)

	Ingested Data			Generated Data
Date	**GOES-16 (NOAA BDP S3)**	**S-NPP/J-1/GCOM-W (NSOF I&T)**	**Ancillary Data (NSOF I&T)**	**NDE Proving Ground Generated Products**
12/19/2018	233.47	1296.89	3.17	5683.13
12/20/2018	309.81	1893.79	5.14	2607.64
12/21/2018	333.19	1910.81	4.50	7563.10
12/22/2018	264.96	1754.37	4.23	4652.45
12/23/2018	264.48	1780.63	4.32	4687.18
12/24/2018	264.71	1756.60	4.31	4944.36
12/25/2018	266.21	1768.81	4.27	4873.24
12/26/2018	179.62	1808.10	4.28	4880.65
12/27/2018	270.44	1766.06	4.30	4896.78
12/28/2018	270.51	1791.80	4.36	4626.84
12/29/2018	270.32	1757.24	4.43	4202.82
12/30/2018	270.68	1726.00	4.38	3776.65
12/31/2018	266.77	1822.01	4.44	6379.45
Average	266.55	1756.39	4.32	4905.72
Median	266.77	1768.81	4.32	4873.24

Analysis and Evaluation Phase

Objectives:

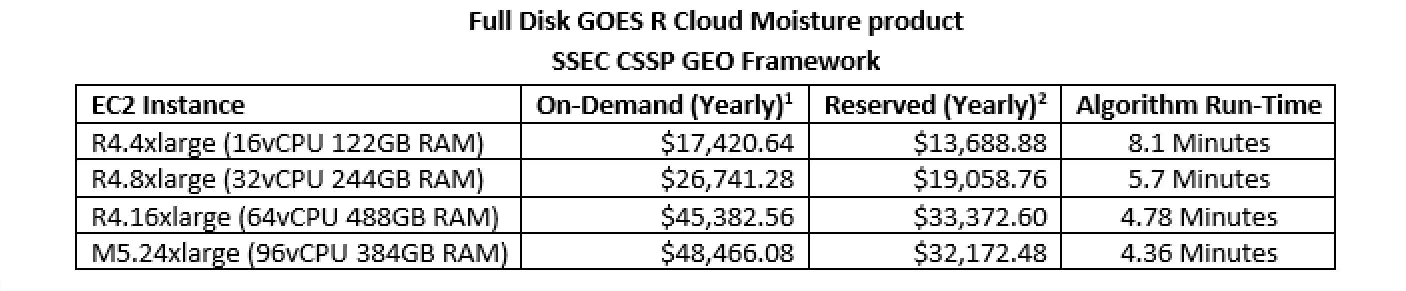
- Provide assessment of ESPDS functionality cost reduction opportunities
- Generate actionable data to support cloud transition cost / benefit decisions
- Support capability transition prioritization
- Provide Enterprise-Ready Data Management and Product Generation cloud capabilities that can support existing "as-is" algorithms, as well as, planned "to-be" modifications. (i.e., containerization)
- Establish baseline of data management/discovery and product generation capabilities for comparison with other cloud prototyping efforts



Analysis and Evaluation Report

Analysis and Evaluation Report Content:

- Cost drivers: ingress, egress, transactions, CPU, services, latency, availability, FISMA compliance, staff, etc.
- Tiered Latency – KPPS vs critical vs best effort
- Single environment vs environment per mission pros/cons
- Cloud provider services vs open source decisions / costs impacts
- Enhancement to FISMA ‘High’ estimates
- TCO of On-Premises vs. Cloud estimates
- Algorithm executable versus containerized algorithm cost/performance comparisons
- Cost per product/latency estimation



Full Disk GOES R Cloud Moisture product
SSEC CSSP GEO Framework

EC2 Instance	On-Demand (Yearly)[1]	Reserved (Yearly)[2]	Algorithm Run-Time
R4.4xlarge (16vCPU 122GB RAM)	$17,420.64	$13,688.88	8.1 Minutes
R4.8xlarge (32vCPU 244GB RAM)	$26,741.28	$19,058.76	5.7 Minutes
R4.16xlarge (64vCPU 488GB RAM)	$45,382.56	$33,372.60	4.78 Minutes
M5.24xlarge (96vCPU 384GB RAM)	$48,466.08	$32,172.48	4.36 Minutes

Questions

SOLERS
An Employee Owned Company