Group hierarchy	Region	Description	Service	Upfront	Monthly	First 12 mc	Currency
My Estimate	US East (Ohio)		Amazon EC2	0	134.1	1609.2	USD
My Estimate	US East (Ohio)		AWS Lambda	0	0	0	USD
My Estimate	US East (Ohio)		S3 Standard	0	117.76	1413.12	USD
My Estimate	US East (Ohio)		Data Transfer	0	92.07	1104.84	USD
My Estimate	US East (Ohio)		AWS Fargate	0	254.3	3051.6	USD

Acknowledgement

* AWS Pricing Calculator provides only an estimate of your AWS fees and doesn't include any taxes that migh

Configuration summary
Operating system (Linux), Quantity (1), Pricing strategy (On-Demand Instances), Storage amount (100 GB), Instanc
Number of requests (1000000)
S3 Standard storage (5 TB per month)
DT Inbound: Not selected (0 TB per month), DT Outbound: Internet (1 TB per month), DT Outbound: Not selected
Average duration (20 minutes), Number of tasks or pods (100 per day), Amount of memory allocated (20 GB), Am

ıt apply. Your actual fees depend on a variety of factors, including your actual usage of AWS services.

:e type (c5.xlarge)

(0 TB per month)
ount of ephemeral storage allocated for Amazon ECS (20 GB)

Project Details

Name of project:

Exploring Cloud backends for the new OGC Environmental Retrieval API

Project lead and contact details:

Steve Olson, National Weather Service, EDMC Deputy Chair (steve.r.olson@noaa.gov)

Project partners and contact details:

Shane Mill, NOAA Affiliate, (shane.mill@noaa.gov)

Proposed start and end date: 11/1/2021 - 5/1/2022

Project Outline

Project description: The Environmental Data Retrieval (EDR) is an Open Geospatial Consortium (OGC) RESTful API standard that makes it easier for users and clients to access a wide range of data through a uniform, well-defined, simple web interface; and provides just the data needed by the user or client while hiding the underlying complexity of contemporary data storage techniques. The EDR's web accessible structured endpoints allow users to extract different feature types from large geo datasets. The NWS has implemented EDR in an AWS cloud environment and is currently conducting a number of prototype studies designed to test the feasibility and impact authoritative web standards, like EDR, have on enhancing both operational and development web applications. A couple of the prototype studies involve connecting to and allowing for feature extractions from AWS Big Data Project (BDP) datasets. These include extracting a time series of streamflow for a few rivers from several TBs of remote National Water Model data and feature extractions of high resolution Himawari and GOES satellite data. The NWS EDR implementation also comes with a user interface (UI), and that allows users to rapidly visualize a request before actually retrieving data for a request. This is particularly important for the Himawari and GOES prototype study, using satellite data available through the NOAA BDP. The NWS EDR implementation uses an AWS EC2 framework and makes use of a Pangeo stack (e.g. Xarray + Dask) on the backend. However, we would like to explore more scalable solutions, certainly serverless AWS Lambda, but potentially also AWS Kubernetes (EKS) and AWS Container Service (ECS). We proposed to experiment with different types of scalable cloud computing to power the Pangeo-based backend we have developed

for the OGC Environmental Data Retrieval (EDR) API.

Project technical & learning objectives:
We have some modest experience with Cloud and AWS (e.g. deploying EC2 instances) but we want to learn more about more scalable and cost effective solutions.

Project significance & impact:
The EDR has the potential to enhance existing community standard APIs delivered by THREDDS and ERDDAP with an international standard developed by the community. Using the Pangeo stack on the backend allows scalability and access to NetCDF4, HDF5, GRIB2, geotiff, COG, and virtual datasets enabled by referenceFileSystem.

Description of key project steps and timeline:
November 2021: Replicate existing EC2 deployment under AWS ESIP account
December 2021 - February 2022: Experiment with serverless compute Lambda as a backend (and other AWS services if necessary)
March 2022: Develop example Jupyter notebooks, document findings on GitHub
April 2022: Create plan for deployment on NOAA cloud resources.

Outreach

What groups will be engaged in the project?

For the National Water Model prototype, we will initially engage with NOAA employees at the National Water Center and researchers interested in cloud workflows at USGS.

For the High Resolution Satellite prototype, we will engage with NOAA/NESDIS employees.

Description of *who* (agencies/individuals) should be aware of this project, i.e. potential outreach targets:

For the National Water model prototype:
Fernando Salas, NOAA, National Water Center: fernando.salas@noaa.gov
Dave Blodgett, USGS, Integrated Modeling, Water Mission Area: dblodgett@usgs.gov

For the High Resolution Satellite Data Prototype in the AWS Cloud:

Dana Ostrenga, NESDIS, Dana.Ostrenga@noaa.gov

Project Partners (as applicable)

Description of project partners (agencies/individuals) and their involvement:

For the National Water model prototype:

Rich Signell, USGS, Coastal and Marine Program, Hazards Mission Area: rsignell@usgs.gov

Will act as advisor on the Pangeo stack and cloud-optimized data

For the High Resolution Satellite Data Prototype in the AWS Cloud:

Dana Ostrenga, NESDIS, Assistant Chief Data Officer: Dana.Ostrenga@noaa.gov

Serves as lead for USGEO NOAA-NASA Satellite Initiative and will act as advisor on cloud-optimized data and retrieval