Migrating NASA Mission Processing to the Cloud: The HelioSwarm Trade Study

Elisabeth Drakatos
Space Science Center, University of New Hampshire, Durham, NH 03824

Introduction

Inefficiencies and modernization initiatives led the UNH Space Science Center to consider processing and storing NASA HelioSwarm Mission data in the Cloud. This allows for a more updated and accessible structure that removes certain on-site storage costs. In developing this plan, Amazon Web Services was the selected Cloud platform to host the processing on. In this study, many tools on Amazon Web Services were tested including AWS Batch for processing, AWS EC2 for Cloud compute capacity, and AWS EFS for Cloud storage. In addition to testing the functionality of the AWS tools, research into pricing strategies and HelioSwarm specific costs for each service was performed. Piecing together a set of tools that worked with the AWS framework and the HelioSwarm Mission's needs was the main objective as well as the main challenge.

Methods

1. In order to narrow down options, processing frequency capabilities and storage features offered by AWS were looked at. During this, pricing research was necessary because many services offered and charged more than was needed for the project.

2. Once a service was selected, it was tested on the AWS website and accessed through Ubuntu. For AWS EC2 instances, compatibility was determined based on the architecture of the instance, its storage, and pricing options.

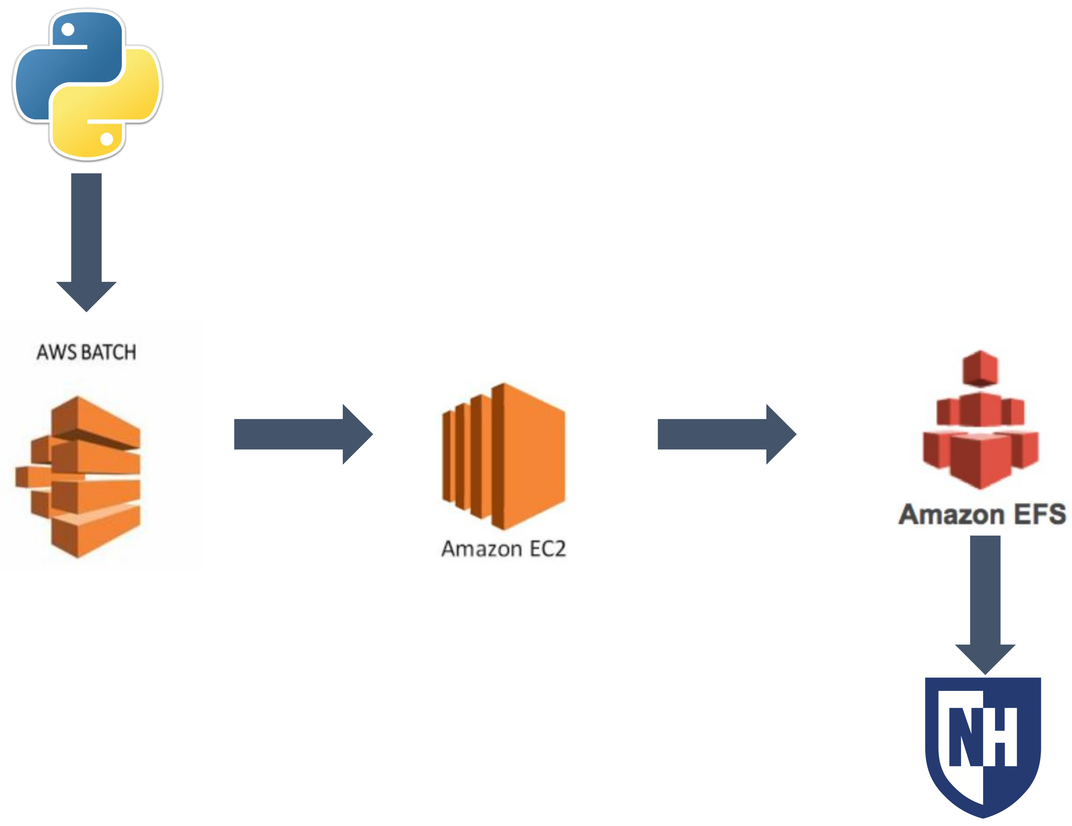
3. Trials on Ubuntu involved mounting an AWS EFS filesystem onto an EC2 instance and migrating data onto it.

4. Investigations into customized Amazon Machine Images took place to assure spin off instances had the necessary software and data downloaded onto them for processing.

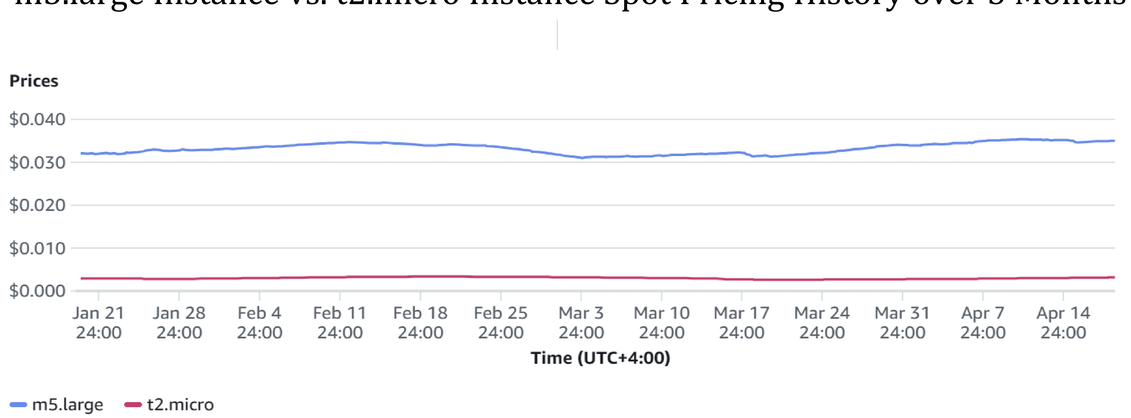
5. Then, a Python application was developed that makes calls to Amazon Web Services using provided request syntax. Specific functions include creating a compute environment, making a job queue, and getting machine images.

6. When our application runs, it launches an AWS Batch job for processing. On the AWS website, information about this job can be viewed such as run-time, instance number, and pricing.

Process



Chart



Results

The results from our study found that the combination of using AWS Batch for processing, AWS EC2 for compute capacity, and AWS EFS for storage worked the best. A main factor in the structure of this combination is due to the opportunity to do AWS Spot Instance Pricing. AWS Spot Instance Pricing allows you to limit the price that you pay by only running the process when the price of the instance is below your pre-determined price limit. Due to the timing flexibility of the data processing, Spot Instance Pricing allowed costs to be less than on-site. For Amazon Machine Images, there were troubles creating one that worked with the AWS Batch processing, so for testing purposes we had to result to using a pre-made AWS one. In terms of storage, both AWS S3 and AWS EFS were explored but EFS ended up being the simpler option.

Conclusions

Our research for this project led us to a solid structure that we will continue to build upon and integrate into the HelioSwarm Mission's needs. In order to progress further, we need to have our AWS Batch process integrate HelioSwarm specific processing requirements. We also need to further set up our EC2 instance so that it has the needed software to spin off during the processing. A main issue that occurred was the AWS Batch process being stuck in runnable. This means that the process was verified to run, but it did not end up running due to incompatibility. This was difficult to solve because the incompatibilities were not listed. In order to subside the runnable issue, we used less customization in our setup.

Next Steps

- Configure a Docker image that is compatible with Ubuntu and AWS ECS.
- Run an AWS Batch process that works with data similar to the HelioSwarm Mission's data.
- Advance our current codebase that accesses the AWS tools programmatically.

Acknowledgements

Special thanks to:

- Dr. Jonathan Niehof and Alana Burch
- The UNH Space Science Center