



AirFlow Processing Pipeline

Mini-Introduction

October 14th, 2022

Who We Are

The Team

- Chris McCorkle
- Isaiah Raspet
- Richard McCormick
- Quinton Jasper
- Hunter Woodruff

Our Sponsor

Trent Hare
USGS

Our Mentor

Vahid Nikoonejad Fard



The Goal

We hope to assist in the creation of a modernised ISIS3 pipeline to streamline the process of charting planetary surfaces

Our Client's Problem...



Main Problems

- Currently, there are over 300 different modules within ISIS3 which are available. Ten of which are in much needed demand
 - This may lead to situations where researchers might not be able to apply the best tools available to them.
- Researchers cannot save pipelines they want to use
 - This makes it more time consuming for researchers to run the correct pipelines they need.

Stretch Problems

- The interface in use currently, a linux terminal, is outdated and unintuitive
 - This may make the learning curve for the software steeper, increasing training time
- The software is currently limited to local use only
 - By allowing cloud-optimized infrastructure, the product and software itself will be much more accessible by the research team

Our Client's Problem (Cont.)



ISIS3

ISIS3 is a tool created by USGS and NASA for allowing scientists and researchers to place different types of cartographic data onto the correct corresponding location of extraterrestrial bodies.

This tool allows scientists to turn raw data collected from remote sensing platforms into actionable, analysis-ready products, such as archives, topographic or cartographic maps, and digital elevation models.

Pipeline Software

Apache Airflow is a pipeline development tool which allows researchers to sequentially process RAW planetary surface data and generate surface models.



These tools have been used to support the safe landings of various extraterrestrial expeditions, such as Perseverance Rover in 2021.

Where We Fit In



The current tools available for this type of data processing can be challenging to use for researchers who are not familiar with computing science. Our goal is to simplify the process so that researchers:

- Have easier access to computational geographic tools for those who need them. We are currently focusing on the airflow processing pipeline, although this may change.
- Can easily create analysis of Martian geographical features
- Can access a continuation of the process tools for mapping the Martian surface, as well as potentially other rocky celestial bodies.

Our Proposed Solution



- Through the use of the ‘Airflow processing pipeline software’, we propose a system of python-wrapped ISIS3 ‘Nodes’ which will allow the scientists at USGS to accurately, efficiently, and quickly create actionable products.
- Python-based module wrappers would allow for the creation of pipelines that can be automated using the AirFlow system.
- Pre-built pipelines can also be built to allow for frequently used configurations to be easily accessed.

Plan for Development



- Requirements Acquisition
 - Continuous client communication via team Discord
 - Issue tracking via GitHub
- Technical Challenges
 - Client requested Apache Airflow
 - Automated wrapping of modules
 - Use of Python as scientific standard
 - Keep application accessible to research team
- Stretch goals:
 - Wrap over 300 modules into the pipeline
 - Create “Recipes” that scientists can use as defaults for specific outputs
 - Push results of pipeline to web service for easier viewing experience

Summary



- Project
 - Development of pipeline software to simplify, save and replicate researchers' computational workflows
- Problem
 - Requiring the use of over 300 individual program modules
 - The USGS team requires a streamlined solution
- Solution
 - Apache AirFlow API
 - Wrap modules into intuitive UI, allowing for ease of use & pipeline reproducibility