

2016 NCSA SPIN INTERN

Developing the TerraRef Plantcv Image Extractor in the Clowder Data Management System

Fengling Wang

Abstract

I worked on writing the TerraRef Plantcv Image extractor in the Clowder Data Management system and the development is under the docker environment of Clowder, RabbitMQ, and MongoDB.

Introduction

My research area is about The Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA-REF). And its major topic is Advanced Field Crop Analytics in Precision Agriculture. Our goal is to accelerate breeding and the commercial release of high-yield bioenergy sorghum hybrids. One of the challenges is the data management challenge - big data volume and velocity.

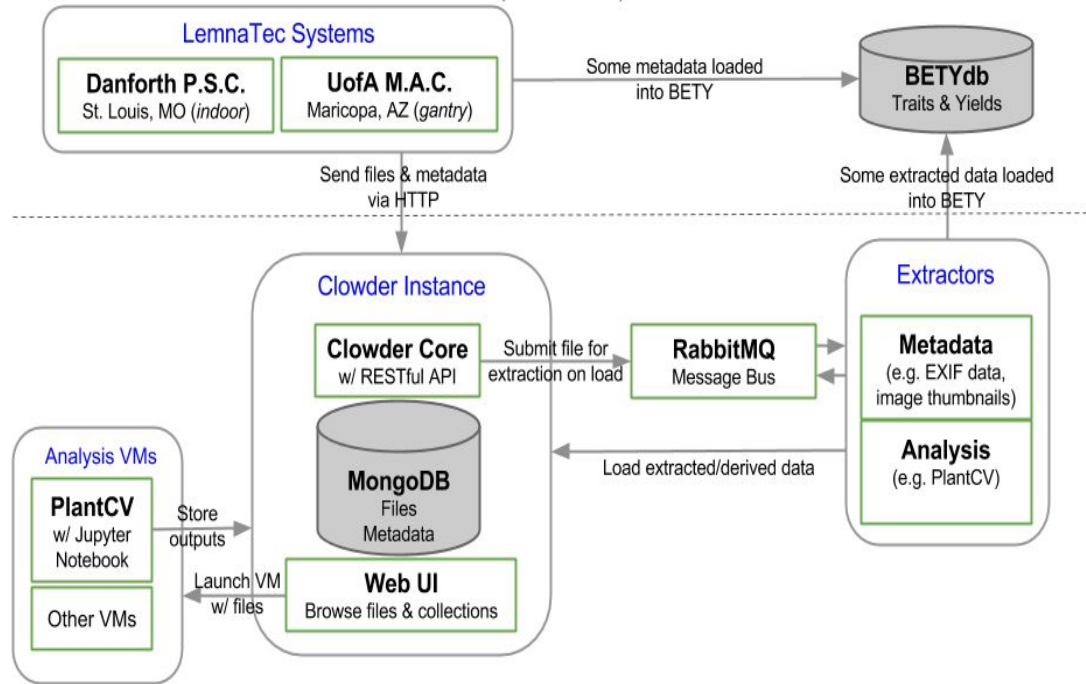
Overview

The overall workflow of the extraction is as below:

1. Upload data to Clowder
 2. Clowder runs on Plantcv Pipeline as an "extractor"
 3. User can select images to view results
 4. User can launch shell or Jupyter notebook with data for future analysis
-

TERRA-REF Pipeline for Data Management and Analysis

(rev. Jan 2016)



Results

Together with other group members, we have successfully finished the first version of the plantcv extractor, which can be run on the ROGER HPC for large batch processing of indoor snapshots. And yet more features are yet to be added and we need better algorithms to make the extractor more verbose.

Discussion

TERRA Ref will provide an unprecedented open-access source of data and an integrated phenotyping system for energy Sorghum. The TERRA Ref system will include field- and controlled-environment digital sensing of energy Sorghum along with computational pipelines and open data for the research community. These will be used for crop selection and better understanding of the interactions among genes, traits, and the environment.

Conclusion

During the development, I've learned how to set up the docker environment of Clowder, RabbitMQ, and MongoDB. I've also learned interfacing the PyClowder API in Python to develop the extractor, and knowledge about modulating and testing the code.

References

<https://github.com/terraref/computing-pipeline/issues/119>

<https://github.com/terraref>

<http://terraref.org/>