## Overview

Cloud Computing Data Processing (CCDP) is an IRAD whose intention is to create a framework to easily interact with the most common services offered by the cloud provider for data processing. The framework takes the burden away from knowledge about the cloud provider from the processing modules. It aims to facilitate an environment that is dynamically modified be either allocating or deallocating resources as needed. This dynamic environment will allow maximizing the resources usage without compromising processing threads. A processing thread is a sequence of processing modules used to generate some results. The framework provides a way to allow communications between the modules.

CCDP can be divided into three main components:

### Command and Control of the Modules

This represents the program whose main function is to assign tasks to the registered processing modules. In order to achieve this we will be using Celery as the tasking queue. Each module should contain some basic configuration such as min and max number of modules required for each processing thread. The framework would allocate as many modules in a single VM as permissible to expedite processing unless the module has a specific function that could be satisfied by one of the cloud services.

### API at the module level

The modules are the components used to process data. The term “process data” is open to each system’s need. Regardless of the processing each module will need a way to provide input and output data. The framework will take advantage of some of the services such as DataPipeline to accommodate the modules in this regards. Besides

### Tools and API used to talk to the AWS services