# CrowdReason Engrafa Document Processor Assembly Specification

## Overview

Currently, the Engrafa Service exists as the Workflow Manager. At high level, the service pulls documents from cloud storage ([Box](https://www.box.com)) and moves them through a series of tasks on a cloud work platform ([CrowdFlower](http://www.crowdflower.com)) to crowd source data extraction. Once each task in the flow has completed, the Workflow Manager aggregates the results into a single structured data file (XML), back to the cloud storage.

The Document Managerwill facilitate uploads and downloads of documents to and from the cloud storage. This responsibility will no longer be handled by the Workflow Manager. Additionally, the Document Manager will use a service ([CVISION Trapeze](http://www.cvisiontech.com/trapeze/general/trapeze.html?lang=eng)) to compress, carry out skew correction, and improve OCR for the documents. Document Manager will leverage Box’s View API to convert PDFs to HTML5 for additional viewing flexibility.

This initial version will be headless (no GUI), but may be extended later to add a frontend.

This assembly will implement the Document Processor and Document Processor Scaling logic in Document Manager.

## Scope

*Note: Please read the whole Application Design Specification first. All the details not mentioned in this specification are provided in that document.*

This assembly will implement the following:

* Models Class Diagram
  + All Mongoose Schema Objects
* Document Processor Class Diagram
  + All classes
* Document Manager Class Diagram
  + config
  + DocumentService.scaleDocumentProcessors

## Deliverables

* Source Code
* Deployment Guide

## Technology overview

* JavaScript
* [Node.js 0.10.29](http://nodejs.org)
* [Express 4.9.7](http://expressjs.com)
* [MongoDB 2.6](http://www.mongodb.org)
* [Mongoose.js 3.8.12](https://www.npmjs.org/package/mongoose)
* [winston 0.7.3](https://www.npmjs.org/package/winston)
* [superagent 0.18.1](https://www.npmjs.org/package/superagent)
* [underscore 1.6.0](https://www.npmjs.org/package/underscore)
* [forever 0.11.1](https://www.npmjs.org/package/forever)
* [aws-sdk 2.0.19](https://www.npmjs.org/package/aws-sdk)
* [shelljs 0.3.0](https://www.npmjs.org/package/shelljs)
* [uuid 2.0.1](https://www.npmjs.org/package/uuid)
* [async 0.9.0](https://www.npmjs.org/package/async)
* [Box Content API](https://developers.box.com/docs/)
* [Box View API](https://developers.box.com/view/)
* [Amazon Web Services](http://aws.amazon.com)
* [Amazon EC2](http://aws.amazon.com/ec2/)
* [CVISION Trapeze](http://www.cvisiontech.com/trapeze/general/trapeze.html?lang=eng)

## Existing Documents

* Class Diagrams
* Sequence Diagrams
* Application Design Specification
* Assembly Specifications