AWS Project Statement of Work

## Objective

User needs help configuring asynchronous jobs to extract text from 30 page pdf files and outputting results of text OCR to csv files.

## Inputs

* Revenue Statements (see attached sample) in pdf format.
* Each statement typically runs from 20-35 pages.
* Total number of statements to process initially, about 50 reports. Thereafter, about two per month.
* The owners, check numbers, dates, amounts, and per-property information will all vary from month to month.
* However, the general layout of the Revenue Statements will be the same.
* Each pdf input file will have a different filename.

## Outputs

* For each pdf file, we wish to get back one text file in csv format (see attached sample).
* The sample shows output that comes from Textract synchronous processing on AWS. This is the desired output. We would like each csv file to look the same way after asynchronous processing with Textract.
* Each output csv file should have a different filename, just as each input pdf file has a different filename. Ideally, the output csv filenames will have enough similarity to the corresponding input filenames so we know which input file generated which output file.

## Processing and Other Requirements

* We would like to automatically trigger Textract processing by uploading files from user’s computer and dropping them into a designated S3 bucket. Files may arrive in the S3 bucket one at a time, or several at a time.
* Output csv files should arrive in a designated S3 bucket where they can be manually downloaded back to a user’s computer.
* Notifications can arrive by any means agreed upon between the programmer and user.
* We don’t care whether the jobs are programmed in Python, PHP, or another language.
* Jobs should automatically clean up resources after completion to eliminate unnecessary charges.
* Jobs should be set up to run under a single authorized user, which programmer and user will set up as part of this engagement.
* Security parameters should be set up to ensure no unauthorized usage of the AWS account or resources once development work is complete.
* Work includes training the user to initiate jobs, monitor status, and receive outputs.
* Work also includes support for troubleshooting or questions during the first 30 days of use.
* Programmer must agree to non-disclosure of all data, programs, sign-on information, and other information shared by users in connection with the project.
* User will own all code generated in connection with this project. However, user will not prohibit the reuse of the same or similar code by the programmer on other projects, at the programmer’s discretion and without compensation to the user.