

Streamlining tax-filing with document OCR and Machine Learning

We designed and implemented a streamlined data architecture, ready to receive and store processed tax documents into BigQuery and enabled data lake insights to be shared with financial partners in a safe and compliant manner.

The challenge

The challenge was to develop a production pipeline that would process a high volume of files, extract with accuracy through accurate ML models, and monitor the entire process end-to-end. Careful consideration had to be taken to manage the asynchronous processes required to upload tax documents, perform data cleansing, process with Optical Character Recognition ("OCR"), categorize and store.

The solution

We began the process to extract valuable data from various document types using existing and newly built parsers, processing for image and data quality, cleaning, and organization of data. Next, we built an architecture to upload, process, and store the extracted data. This reliably synchronized the workloads that run through custom pipelines, with continuous learning and monitoring.

The results

By implementing machine learning tasks like intelligent document character recognition, state-of-the-art features were able to be launched for consumers and serve as a source of valuable client insights for professionals in the financial industry. Data Loss prevention and security safeguards were implemented to ensure anonymized data to share with partners and the Machine Learning applications.

1. Event-driven architecture enabled ingestion over 100,000 documents per hour and runs Optical Character Recognition at 10,000 documents per hour
2. Human in the loop data verification and automation to save time to ingest, verify, and parse data, down from 3 hours to just 15 minutes
3. End-to-end deployment of new OCR tax parsers under 10 minutes
4. Real-time data quality assessment in place

About Halcyon Solutions Corp

Halcyon Stillwater offers a portal for clients to securely access financial information and digitally prepare tax returns approved by tax professionals. They are also innovating ways for financial professionals to access valuable client insights.

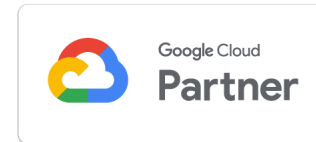
Industry: Financial Services

Primary project location: Canada



About Bitstrapped Inc

Bitstrapped is a Google Cloud Partner specialized in the training and transformation of businesses through Data and Machine Learning platforms



Products

Google Cloud Platform