

Contact

862-336-0434 (Mobile)
henryclavo@gmail.com

www.linkedin.com/in/
henryorlandoclavo (LinkedIn)

Top Skills

Claude Agent SDK
Azure Functions
Scikit-Learn

Languages

English (Native or Bilingual)
Spanish (Native or Bilingual)

Certifications

Functional Programing in Python
Lean Six Sigma Yellow Belt
Certification
Speed Up Python With Concurrency

Understanding Python List
Comprehensions

Honors-Awards

Honor Society
Verisk Way to Go

Publications

Guidewire Accelerator Guide

Henry Orlando Clavo

Technology Consultant | Data Scientist | Big Data Engineer | Cloud Engineer | AI Autonomous Agent Development | Data Strategy | Databricks & Microsoft Fabric | Speaker | AI Consultant | AI Engineer
United States

Summary

Passionate and energetic data professional who loves a challenge. I am interested in intellectual work with rigour. I look for opportunities that offers unique perspectives on social and economic issues. I consult for companies in delivering results for their business using theoretical and mathematical concepts.

I have implemented technology solutions for government agencies, healthcare hospitals, financial services, and Insurance companies (16 clients). Able to efficiently solve the customer's ETL project issues. Work with technical and non technical staff to provide clarity and solutions to customer's needs in technology.

I enjoy working with people and helping them achieve their potential to reach success. I build partnerships and relationships with my clients, and work with them to meet their needs. I work with different businesses and IT development teams to work on technical and operational improvements that lead to innovation.

Indoor and beach volleyball player when not doing data science in my free time.

Specialties:

- Data Science (Natural Language Processing, Regular Expression text extraction)
- Data Engineering (AWS cloud implementation)
- Product Management
- Writer Contributor (Guidewire manual)
- Python Developer
- Customer Implementation
- Configuration Management and Cloud Lead

Program Languages/Technologies:

- R
- Python (Pandas, Spacy, Great Expectations, Boto3)
- Pyspark
- Docker
- SQL (Mysql, Microsoft SQL Server Management Studio, DB2, TablePlus, Postgres)
- Cobol/JCL
- Mainframe
- Linux (Ubuntu)
- Go Lang
- Stanford Core NLP
- AWS (S3, RDS, Glue, API Gateway, Lambda function)
- Azure Devops (Data Pipeline, Git)
- Tableau (Stacked Bar Chart, Heat map, Dual Axis Graph, Custom Calculations)
- Tesseract OCR & Google Cloud Vision API
- Salesforce (SOQL Query and REST API)
- Puppet and Ansible

Accomplishments:

2020 CFPB: Technology implementation & Concurrency

2020 Project Covid-19: Worked with Data Governance lead to build ETL solutions for government agencies and develop the product with the engineering team. Developed scripts alongside engineering team for automation process.

2018 Guidewire Accelerator: Working with Development and Business Analysts in testing and analyzing data for this new product to help insurers convert data easily from their system.

2015 Mastercard Consumer and Small Business Study: The study resulted in expansion of the work in 2016 to cover South American and Canadian markets. Understanding credit card usage from consumers and small business owners.

For speaking and podcasts opportunities, reach out here.

Experience

Consumer Financial Protection Bureau

Data Engineer

May 2020 - Present (6 years 3 months)

United States

Supporting the agency with modernization efforts and technology implementation. Implementation of a new technology that added new features for release in 2021. A joint effort with Consumer Response and Technology & Innovation. Daily use of postgres, docker, SQL, Python, AWS(EMR,S3,EC2), and Elasticsearch.

2021 projects:

- Provide support to Complaint Analytics.
- Developing docker environments and settings for testing pipelines.
- AWS development POC (stream and batch processing) using EMR and open source D.Q.
- Provide support to cross functional teams.
- Create ETL process to manage metadata for all data pipelines used and create data model.
- Experience with AWS EMR, Kinesis, API Gateway, S3, Glue, cloudwatch, and lambda.
- Building docker containers using YAML and images used PGADMIN, Postgres, Elasticsearch, and Kibana.

Successful outcomes 2021:

- Provided product's continued success in development by processing 7.5 million records in search.
- Implemented concurrency process for pipeline completion and increase efficiency by 50%.
- Reduced OCR data processing time by 50%. More than 5,000 data points processed daily.
- Added new features to display statistics for data pipeline processing.
- Modernized tech stack by refactoring data pipelines to improve performance and leverage new technologies.
- Implemented Python data processing from database to Elasticsearch api for CFPB internal tool (400,000+ data points processed in July).
- Worked with developer to provide business requirements to correct 200,000+ data points affecting review of process.

-Created workflows and roadmap for organization to build many data sources from data lakes to data bricks environment.

M9 Solutions

Python Data Engineer

May 2020 - October 2024 (4 years 6 months)

United States

See CFPB

Accenture Federal Services

5 months

Data Scientist | Applied Intelligence

January 2020 - May 2020 (5 months)

Arlington, Virginia, United States

Discovery Lab Product Development:

Projects

- Building a data collection automation process for management.
- Use of Natural Language Processing to extract and analyze text data.
- Used Spacy package to extract text data from sources for a Machine Learning model.
- Work with Program Manager to test use cases and scenarios for the product.
- Develop automation process using Python and SQL tools.
- Work with team members in project to polish and optimize pipeline code.
- Develop data pipelines using python from different data collection sources.
- Web scrape and develop data collection scripts for automation using BeautifulSoup and Selenium.

Operations Lead:

- Promoted to lead AWS project efforts and managing two Data Engineers.
- Build Data Pipeline using python scripts
- Work with team to set up APIs for scalability (Amazon API Gateway)
- Lead the Engineering Group
- Attend Data Science modeling group for data support
- Provide updates to leadership and sales
- Work with Data Governance lead to build database standard

Operations Lead (Remote) | Applied Intelligence

January 2020 - May 2020 (5 months)

Arlington, Virginia, United States

Applied Intelligence

Powered by Management Consulting and Technology, AFS Applied Intelligence (AIX) is a full-service data-native practice group that helps federal departments and agencies unlock business value from their data, improve visibility and performance across the enterprise, enhance decision-support and mission outcomes, and accelerate digital capabilities at scale.

We are at the forefront of helping federal agencies apply more innovative approaches to achieve their mission objectives and improve operations. As leaders in advanced analytics, artificial intelligence (AI), and automation, our cross-disciplinary team of data science, data engineering and data visualization strategists and practitioners provide deep technical, functional and mission expertise, innovative solutions, and tested methodologies that help agencies collect and measure what matters to them, intelligently harness emerging technologies, optimize decision-making, and holistically and responsibly drive end-to-end change.

Capabilities

Our offerings for the Federal sector are designed to support customers at every phase of their data journey:

Artificial Intelligence and Automation

We help our clients establish strong practices and executable roadmaps that set the stage for any AI or intelligent automation transformation. Our offering includes AI Strategy, AI Systems and Service Design, Responsible AI, Conversational AI, and Robotic Process Automation.

ASSETS AND ACCELERATORS

We support portfolios and our AFS clients through multiple channels:

The Discovery Lab offers prototyping, strategy, and design services across the full spectrum of Applied Intelligence offerings. Co-located at our Digital Studio in Washington, D.C., the Discovery Lab is our hub for data discovery, innovation, and emerging technologies.

Verisk

Business Analyst

November 2017 - November 2019 (2 years 1 month)

Jersey City, New Jersey, United States

Earned Verisk Way to Go Award for excellence in data analysis and research:

issue of misreporting 1,300 policies for Geico from a data load was resolved through data analysis and collaboration with development to resolve any issues to state DMV.

Product Owner:

Guidewire Accelerator. Released Product that supports client's data implementation process.

Coverage Verifier Conversion Project. Released first iteration of program to clean up data preprocessing step.

Data Quality Team :

Analyzed current format of data ingestion process. Communicate with management on process to improve data processing. Capture missing data elements for each customer in QA process. Train team members to report data quality process, business logic, and implementation of new logic for customers. Work with development team and QA team to test data program and business logic. Prioritize and release features for the project in two week sprints in agile.

Migration Consultant:

Worked with ETL group to consult business logic and QA for Coverage Verifier migration from Legacy (Mainframe) to the cloud (AWS). Worked extensively with group during meetings to provide business requirements.

Technology Solutions:

Worked extensively to study Coverage Verifier flow to understand issue of missing reports from the clients. Investigate files transfer protocol from the Mainframe to production SFTP server. Discovered with team of the corrupted files when the network disconnects. Documented and reported incident to management to execute a solution with data team and data quality team.

WeLIFT.co

Lead Data Scientist

2016 - February 2018 (2 years)

New Jersey

As a Data Engineer, my role is to migrate processes from the old system to a new platform. Before the analysis, I need to build the structure and pipes. Once this process is completed, I will be able to generate reports and collect and manage data flow for the organization.

MedAssets

Data Analyst

October 2016 - November 2017 (1 year 2 months)

Greater New York City Area

Client: RWJ Barnabas Health

Identified data quality issue within client's application and used SQL functions to update database records. Leveraged the test environment to quickly test and create a new table and data set to push the update to production. I quickly resolved the issue for Robert Wood Johnson's one of many facilities in New Jersey experiencing inefficiencies in data quality.

New Jersey Citizen Action

Consultant

August 2016 - November 2016 (4 months)

Newark, New Jersey

Data Analyst:

Provided industry and technical knowledge on customer relationship management for a new start up. Provided QA for new data collection system.

Jersey City project:

Manage the organization's project to provide financial services to communities in Jersey City. Consult on best practice on financial services and products for consumers who are in debt. Collect data during interviews and provide an outline to clients looking to improve their finances and decrease spend.

Ipsos

Statistical Analyst

May 2015 - May 2016 (1 year 1 month)

Parsippany, New Jersey

Mastercard 2015 Project:

Analyzed, cleaned, and standardized data for Mastercard's credit card data for the North American Market. Worked with Ipsos Loyalty director, Mastercard consultant, and intern to prepare the data for analysis by using outlier detection method, statistical inference and comparison of means to understand consumers's spending habits and top ten credit card usage in their wallet. The success of Mastercard's consumer and small business study in 2015 resulted in expansion of the study in 2016 to cover South American and Canadian markets.

My Personal Accomplishment:

Teaching my intern on the use of SPSS syntax to automate data cleaning process, and working together to prepare the data for our client.

Kids Corporation II

Research Associate

May 2014 - July 2015 (1 year 3 months)

Newark, New Jersey

- Administered surveys to one hundred students for summer program evaluation
- Created surveys to collect data on the impact of afterschool programs in Newark Public Schools
- Used Microsoft Excel and IBM SPSS tools to create charts and tables for Statistical Analysis
- Created a final report for partners to continue funding Kids Corporation II for Summer 2015

The Financial Clinic

Financial Fellow

July 2013 - August 2014 (1 year 2 months)

- # Managed two sites in New Jersey for direct services and achieved contract obligation
- # Documented and tracked customer data for data manager
- # Provided workshops and presentations for NY/NJ customers
- # Formed partnerships with non-profit organizations throughout New Jersey
- # Invited speaker for the Lift Up project in Savannah, Georgia for services and issues that come with financial hardships

Education

Quantic School of Business and Technology

Executive MBA, Master of Business Administration · (January 2020 - November 2023)

Trivera Technologies

Certification, Applied Python for Data Science · (2020 - 2020)

Montclair State University

Certification, Data Collection and Management · (2016 - 2017)

edX

Certification, Introduction to Python · (2016)

Rutgers University

Bachelor's degree, Majored in Economics with minor Political Science · (2008 - 2013)