TODD MARINO
DATA SCIENTIST/ ENGINEER

CONTACT

+1 (771) 201-8442

marino.todd@gmail.com

www.github.com/marinert
www.linkedin.com/in/toddadammarino

Northern Virginia

SKILLS & TOOLS

Data Engineering Products

Detailed & client architectural diagrams, full pipeline, implementation, RedHat OpenShift, Gitlab Ci/CD, AWS, CloudFormation, Rancher, GitOps

ML/AI Products

Recommenders, NLP, Predictive Analytics, Classifiers, Regressors, Ensemble Methods

Methodologies

AGILE, Waterfall, CI/CD

Datastores

Delta Lakes, Lakehouses, Data Warehouses, NoSQL, SQL, HDFS, Graph Database

Programming Languages

Python, YAML, Bash, Helm, Scala, HTML, JS

EDUCATION

Bachelors of Art, 2001 - 2005
The George Washington University

MBA, 2009 - 2010
Rutgers University

Certificate in Modern Standard Arabic, 2013-2015
Defense Language Institute

Certificate in Data Science, 2019-2020
Galvanize, Inc.

CERTIFICATES

Exp. July 2024
Apache Spark Developer Associate

Exp.Nov 2024
AWS Certified Solutions Architect

Exp. Sep 2026
CompTIA Security+

ICAgile Certificate

PROFILE

Self-motivated, client-oriented, collaborative data engineer/data scientist with 7+ years in IT/Cyber and 4+ years experience in Data Science who enjoys creating, reusable, scalable data pipelines. Frequently translated vague client needs into specific actionable plans, thereby resulting in improved, efficient, processes with more reliable data. Moved an entire team to distributed processing with custom Python packages and detailed diagrams for continuity and documentation purposes.

WORK EXPERIENCE

Senior Data Scientist

Booz Allen Hamilton   2021 - 2023

- Lead an Artificial Intelligence project researching ways to utilize Large Language Models with LangChain and Hugging Face Hub tools like its transformers.
- Established benchmarks, tasks and statistical KPIs for LLM evaluation.
- Develop and manage end-to-end data from collection of GDELT to deployment through a Rancher managed Kubernetes cluster to a public Tableau Server, as well as create Tableau dashboards.
- Collaborated on prototypes for different sources (web scrape, api, SQL, NOSQL) for analytics products including NOTAMS and ITU Satellite APIs.
- Developed geospatial visualizations in Tableau and Streamlit for a variety of products including heatmaps and locations.
- Improved temporal-geospatial data analysis timing by 324% over a 6 month period, improving concision and tripling geospatial coverage for analysis from 6 to 20 countries.
- Collaborated with team members to eliminate data siloing by creating Databricks-specific patches to existing packages and developing re-usable, internal python packages.
- Streamlined and automated analytical modeling from data preprocessing to model evaluation while evaluating the statistical effects of the data mining changes.
- Performed ETL operations on API accessed data, creating continuous updated raw data sources stored in data lakes using Databricks Lakehouse infrastructure from external sources.
- Pitched, developed, and implemented an entire Apache Structured Streaming ETL process incorporating internal Python packages to transform and clean lakehouse stored data.
- Mentored new data engineers engaged in a company sponsored bootcamp, advising on company culture, company resources, and answering questions about data engineering resources .

Data Scientist

Booz Allen Hamilton   2021-2021

- Collaborated on a cross-functional, scrum-based team focused on AGILE software development to turn manual analysis into automated analytics with SpaCy's Named Entity Recognition ruler algorithm.
- Generated test data and experiments in absence of datasets.
- Worked closely with other team members to ensure proper model implementation.
- Customized a SpaCy 3.0 natural language processing CNN model with rule-based matching focusing on creating unique labels for capturing data in messages.
- Built, tested, and deployed a REST API with FastAPI on AWS Cloud infrastructure and Docker using Pydantic modules for data consumption and data validation.
- Interpreted architect requirements into code, negotiating features with the front-end developer and taking ownership of API and database software components and testing.

TODD MARINO

DATA ENGINEER

CONTACT

📱 +1 (771) 201-8442

✉ marino.todd@gmail.com

🌐 www.github.com/marinert
www.linkedin.com/in/toddadammarino

📍 Washington, DC

PROJECTS

- **Resume Chatbot.** Created a LLM chatbot that interacts with my PDF resume. toddsresumebot.streamlit.app
- **Document Recommender.** Trained and developed a model that recommends documents from EDGAR based on IP address.
- **Message Reader.** Created an NLP model using SpaCy that read messages and labeled entities.
- **ETL pipeline with Distributed Processing using CI/CD and Delta Lake**.
- **Lean ML Pipeline.** Developed a containerized YOLOV5 object detection model pipeline under 13 MB in total.
- **Open Source Analytic Product**. Deployed an app gathering up-to-date data displayed in Streamlit on a Rancher-K8 cluster.

COMPETENCIES

- Linguistics (12 yrs)
- Machine Learning (3 yrs)
- Natural Language Processing (1 yr)
- Presentation (2 yrs)
- Data Engineering/Architecture (4 yrs)
- Quantitative & Qualitative Analysis (5 yrs)
- Curriculum Development & Training (3 yrs)
- IT & Cybersecurity (3 yrs)
- Military Intelligence (7 yrs)

WORK EXPERIENCE

Software Developer/Data Scientist

Buffalo Group — 2020

- Reviewed and updated python script documentation as well as writing a continuity guide using visuals and non-technical language to describe the data ingestion pipeline for collecting, validating, processing, stacking and updating the SQL database.
- Evaluated performance management and formulated recommendations for senior leadership with minimal supervision.
- Explained mathematical formulas and statistical findings to non-technical users and decision makers.
- Utilized Python for data analysis (e.g., Pandas, NumPy, matplotlib, etc...) and statistical significance.
- Mentored junior staff in data governance, coding, statistical modeling.

CyberOps NCO

NJ Army National Guard — 2018-2019

- Awarded Army Commendation Medal for performance and contribution in support of Title 10 operations during Task Force Echo II.
- Initiated and planned section concepts to reorganize and more efficiently utilize existing resource infrastructure.
- Provided KM updates in staff meetings, training courses, online meetings, and other selected venues to ensure the KM program is in concert with ARCYBER knowledge strategy and knowledge requirements.
- Constructed SharePoint pages, web parts, views, lists and libraries to manage content and administer daily operations and user access.
- Planned, executed, monitored, and controlled training for new knowledge managers and knowledge management staff.

Information Technology Specialist

ManTech — 2017-2019

- * Troubleshot daily maintenance on customers' computers, network connectivity, local administrator tasks like creating accounts.
- * Assisted the Applications Team in developing and incorporating Jira and Confluence into the Center's infrastructure.

DOMEX Screener - Arabic

Valbin Corporation — 2016-2017

- Tagged & prepared gists and translations for correct labeling into the TED data repository.
- Demonstrated history of working on screening and translation projects while maintaining the integrity and meaning of the material.
- Scanned and processed a large amount of material in Arabic for critical and essential elements.
- Conducted and prepared accurate gists and translations of general and technical materials.

Cryptologic Linguist/SIGINT - TRAINING

New Jersey Army National Guard — 2012--2019

- Reserve component NSA remote training as part of a Multi-Function Team.
- Field training with the Prophet Spiral 1, including installing, troubleshooting, and maintaining the software and hardware.
- Completed online SIGINT training at the Northeast ARISC including Introduction to Global Networks and Antennas and Radio Wave Propagation.
- Participated in CERDEC Cyberspace Electromagnetic Activities (CEMA) Field-Based Risk Reduction Event (FBRR), an exercise utilizing SIGINT technologies and programs.
- Participated in SSE training using biometric devices, investigative interviewing through the Reid Course and conducting HUMINT mission.