**Todd Marino**

**Booz Allen Hamilton** **Clearance**: Top Secret/ SCI

**Email**: marino\_todd@bah.com **Residence**: Herndon, VA

(771) 201-8442 (cell)

**EXPERIENCE SUMMARY**

Self-motivated, client-oriented, collaborative senior data scientist/engineer with 7+ years in the IT/Cyber realm and 4+ years in data science.

**COMPETENCIES**

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

**SUMMARY OF SKILLS, TOOLS, AND TECHNOLOGIES:**

**Programming Languages:** Python (5 yrs), YAML (2 yrs)

**Python Libraries Familiar with**: PySpark, FastAPI, Pydantic, TensorFlow, Scikitlearn, SciPy, FastAPI, NetworkX, SpaCy, BeautifulSoup, SciPy, PyTorch, LangChain, HuggingFace

**Datastores familiar with:** MongoDB, DynamoDB, AWS S3, Data Lakes, Delta Lakes, Lakehouse, Databricks, SQL, NoSQL, Graph Databases, Relational Databases

**PREVIOUS CUSTOMERS:**

* **DEFENSE INTELLIGENCE AGENCY (DIA)**
* **NATIONAL GEO-SPATIAL AGENCY (NGA)**
* **CYBERCOM, EUCOM, AFRICOM**

**CERTIFICATIONS:**

Apache Spark Developer Associate, Exp. July 2024

AWS Certified Solutions Architect, Exp. Nov 2024

AWS Certified Cloud Practitioner, Exp. Nov 2024

CompTIA Security+, Exp. Sep 2026

Google IT Automation with Python Certificate

DeepLearning.AI Build Basic Generative Adversarial Networks Certificate

DeepLearning.AI Tensorflow Developer Professional Certificate

* 80-hour course with labs and graded quizzes
* NLP & CNN in Tensorflow Certificates

**EDUCATION & TRAINING DETAILS**

Galvanize, Data Science Immersive 12/2019 – 03/2020

San Francisco, CA

* Learned then trained and developed machine learning models including for use in natural language processing, neural networks, regression analysis and classification problems.
* Learned then used statistical and graphical Python packages for analyzing results of the models.
* Generated test data according to different statistical distributions.
* Presented weekly presentations on datasets and model selection criteria for case studies using visualizations generated through seaborn and matplotlib statistical modeling features.
* Provided data-driven insights into the datasets to stakeholders in weekly presentations.
* Designed, developed, analyzed, and deployed individual capstone projects with instructor input and guidance in a scrum-based waterfall development approach.
* Prototyped several models in the final capstone using statistical models and graphical aids for decision making.

New Jersey National Guard, Cryptologic Linguist 12/2012 – 11/2019

Fort Dix, NJ

* 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 tasks.

MBA in Professional Accounting, Rutgers University 05/2009 – 10/2010

Newark, NJ

* Coursework included Information Technology which focused on current technologies used in accounting such as Accounting Information Systems (AIS) to provide data-driven analysis for auditing procedures. The Auditing course incorporated practical experience with Audit Command Language (ACL), a feature of AIS.

BA in International Affairs, The George Washington University 08/2001 – 05/2005

Washington, DC

**EXPERIENCE DETAILS**

Booz Allen Hamilton, Associate 01/2021 – Present

Data Scientist 03/2023 – Present

Chantilly, VA

Project A

* Solely responsible for end-to-end pipeline development, data architecting, and dashboard creation on a self-hosted Tableau server.
* Orchestrated, deployed, and diagramed a Rancher-managed Kubernetes deployment of a Tableau server with CRON job data calls dropping gathered and cleaned data into a MYSQL server with attached EBS volumes.
* Created external and internal ingresses for customer access to the Tableau server’s dashboards.
* Troubleshot issues that arose with Tableau overloading data, making incorrect calls, and showing incorrect data (time-range issues), as well as with the Python client gathering data.

Project B

* Collaborated on a research project to develop a pipeline to deploy an object detection model on a satellite in orbit.
* Leveraged the firm’s international partner, Spiral Blue, to build out and test Docker containers on its proprietary middleware.
* Successfully containerized and orchestrated a lean ML pipeline utilizing under 13 MB of memory in deployment.

Project C

* Using a scrum-based setup, our team developed an analytics product to showcase a variety of different data gathering techniques and visualization capabilities, including web scraping, API calls.
* Showcased a variety of different visualizations using user tables, graphs, and dynamical maps.
* Incorporated data from the International Telecommunications Union (ITU) scraping Network Satellite filings, pulled and cleaned orbital object data from the SpaceTrack API, as well as adding filterable NOTAM data to Streamlit Folium map.
* Worked with the data engineer to debug architectural issues including improving latency and incorporating a REST API.
* Assisted junior staff with coding tips and guidelines for producing NLP products, including using BeautifulSoup to scrape news sites, programmatically update tables on a Streamlit app, as well as use this data to run through a SpaCy named entity recognition model.
* Deployed Streamlit apps to a Rancher-managed Kubernetes cluster.

Project D

* Researched ETL, results validation and deployment methods for large language models.
* Prototyped LLM-chatbots that interact with private repositories.

Data Scientist 11/2022 – 03/2023

Molesworth, UK

* Prototyped modernized CI/CD DevOps pipelines for smoother code exchange, data intakes and enhanced internal processes for managing code.
* Explored different data repositories, processors, and workflow orchestration tools like Jenkins and GitLab.
* Collaborated with colleagues on updating, standardizing Scala code into Python.

Senior Data Scientist 12/2021 – 11/2022

Alexandria, VA

* Improved temporal-geospatial data analysis timing by 324% over a 6-month period using Apache Sedona for geospatial distributed processing and Apache Structured Streaming for accurate reporting as well as tripling coverage from 6 to 20 countries.
* Collaborated with team members to eliminate data siloing by creating Databricks-specific patches to existing packages and code.
* Developed a back to front pipeline prototype for geospatial visualizations using GitLab, Databricks and RedHat OpenShift.
* Designed the MLOps architecture to codify different tables with schemas and use standardized shapefiles.
* 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.
* Developed ETL pipelines for structured, unstructured, and geospatial data, including shapefiles and geojson from the GeoBoundaries API.
* Incorporated Apache Structured Streaming into the pipeline for better error handling, consistency, and fault tolerance.
* Attended weekly scrums and technical exchanges on time series analysis.

Data Scientist 02/2021 – 08/2021

McLean, VA

* 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.
* Codified captured data architecture with Pydantic by creating models that validated data types and extended descriptions.
* Developed the ontology, lexicon and data dictionary for a graph-like database deployed in MongoDB and its interaction with the backend server.
* Interpreted architect requirements into code, negotiating features with the front-end developer and taking ownership of API and database software components and testing.
* Assisted in developing database schema to accommodate future growth into Knowledge Graphs.

The Buffalo Group, Developer 10/2020 – 12/2020

McLean, VA

* Explored the capabilities of TabPy, an open-source library that allows executable python code in Tableau.
* Successfully instantiated AWS Elasticsearch services (v. 5.3), creating mappings and indexing text in its datastore through a serverless deployment using AWS Cloud Formation.
* Evaluated Apache NiFi and Tableau's Prep Builder as possible ETL tools by developing experiments in the absence of training data.

The Buffalo Group, Data Scientist 08/2020 – 10/2020

McLean, VA

* 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.
* Extracted business insights from analysis of data and communicated (orally, written, and visually) those insights to business leaders.
* Utilized Python for data analysis (e.g., Pandas, NumPy, matplotlib, etc...) and statistical significance.
* Mentored junior staff in data governance, coding, statistical modeling.
* Developed recommendations for automating data ingestion and future data analysis.

NJ Army National Guard, Cyber Ops NCO 02/2018 – 04/2019

Fort Meade, MD

* 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.

Valbin Corporation, DOMEX Screener - Arabic 08/2016 – 12/2017

Bethesda, MD

* Tagged & prepared gists and translations for correct labeling into the TED data repository.
* Developed Excel tools to facilitate accurate gisting and translation.
* 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.

Accounting Principals, Analyst 01/2011 – 10/2012 Princeton, NJ

* Mined data from key global financial databases in several languages to identify missing information in Bloomberg’s database.
* Contracted to Bloomberg for two projects, utilized Bloomberg Terminal and related proprietary software to upload, tag and analyze multi-lingual financial documents for about 60,000 periods.
* Researched missing financial periods for global companies in multiple languages, namely in North and West Germanic, West Slavic, and Korean languages, creating standard operating procedures for document retrieval for several foreign markets.
* Thoroughly researched, tagged, cleaned, and manipulated specialized data elements for various dashboards including Blomberg’s Pipeline, Retail Apparel, and Shipping sites.
* Mined data from key global financial databases in several languages to identify missing information in Bloomberg’s database.

The Korea Herald Institute, ESL Instructor 11/2007 – 04/2009

Seoul, South Korea

* Trained and mentored junior staff in how to properly conduct class, foster an encouraging learning environment, and research and develop materials.
* Negotiated, developed, and delivered on-site programs for multinational corporations.
* Demonstrated competency in delivering high-quality content exhibited by a total year-over-year enrollment increase of 15x.
* Responsible for designing, developing, and executing conversation courses for advanced adult learners.