

Stephen Y. Hwang

(224) 730-9044
stephenyounghwang@gmail.com
linkedin.com/in/hwangstephen

Technical certifications

AWS Certified Solutions Architect (Associate) MongoDB Database Administrator MongoDB Developer (Associate)

Skills

Programming languages	Python, R/Shiny, JavaScript, SQL, Perl
Databases	MongoDB, MySQL, Postgres, BigQuery, BigTable, Snowflake, Redshift, EMR
Data engineering	Kafka, Cloud Composer (Airflow), Google Dataproc (Spark), Pub/Sub, SNS/SQS
Data sharing	FastAPI, Flask, Google Looker (Data Studio), Tableau, ggplot, matplotlib, Prometheus
DevOps & MLOps	Git, Docker, Kubernetes, MLflow, CI/CD, Infrastructure-as-code (Terraform, CloudFormation)

Professional Experience

Solutions Architect, Applied AI/ML & Search Research Triangle Institute, RTP, NC Jul 2023 - present

- Architected and deployed enterprise AI/ML applications for semantic classification and data harmonization
- Optimized retrieval-augmented generation pipelines for domain-specific applications using Chain-of-Thought reasoning
- Designed and implemented complex ETL/ELT pipelines to extract business value from proprietary data
- Built scalable FastAPI interfaces to connect modular components in a data platform
- Containerized applications with Docker and orchestrated real-time streaming workflows using Kafka
- Developed hybrid search solutions combining full-text Lucene search with vector search to enhance information retrieval
- Led development of LLM-powered chatbots with memory management in compliance with data security requirements
- Provided technical enablement around hybrid search and DataOps strategies

Enterprise Solutions Architect, Presales MongoDB, New York, NY Jan 2022 - Jul 2023

Partnered with sales teams to support and onboard enterprise customers in healthcare and financial services

- Designed and deployed scalable, event-driven solutions to meet business requirements for digital payments, consumer lending, and the management of electronic health records and biometric data
- Advised customers on best practices for data modeling, distributed computing, and event-driven architectures
- Provided technical enablement, delivered tailored demonstrations, and built proof-of-concepts for technical validation

Senior Bioinformatics Data Scientist Research Triangle Institute, RTP, NC Jun 2015 - Jan 2022

- Designed and implemented a scalable, event-driven cloud framework to automate bioinformatics workflows
- Led cross-functional teams of analysts and developers as a technical lead and program manager
- Developed and optimized ETL pipelines to ingest, process, and integrate multimodal biological and clinical data, enabling reproducibility and improving data accessibility and downstream analytics.
- Built interactive dashboards to enable data-driven decision-making for critical stakeholders
- Translated complex technical concepts into actionable insights for non-technical stakeholders, ensuring alignment between business requirements and bioinformatics solutions.

Solutions Engineer Oncotect, Raleigh, NC Jan 2021 - Jan 2022

- Designed and built event-driven pipelines to extract, transform and ingest experimental biomedical data

Bioinformatics Data Scientist Johns Hopkins Medical Institutions, Baltimore, MD Oct 2011 - Jun 2015

- Lead developer on Stand Up 2 Cancer Epigenetics Dream Team (<https://www.su2c.org/>)
- Developed data pipelines to promote data interoperability across disparate public health and lab systems
- Developed machine learning methods for exploratory analysis
- Developed event-driven workflows to aggregate, process, and enable consumption of public health and genomic data

Education

Johns Hopkins University, Whiting School of Engineering M.S. Bioinformatics, Aug 2013 Baltimore, MD

Loyola University B.S. Bioinformatics, May 2011 Chicago, IL