Tony Kipkemboi

LinkedIn | GitHub | Twitter | ENS: tonykipkemboi.eth

Email : youremail@gmail.com

Mobile : +1 000-000-0000

EXPERIENCE

- **Booz Allen Hamilton** — Arlington, VA
  Data Engineer — *Mar 2022 – Present*
  - **Advana Jupiter Developer**: Supporting PMS 505 project with automating data ingestion from disparate sources (Extract Transform Load) using Databricks and S3 as a Data Lake
  - **Data Engineering**: Completed the AWS Data Engineering Tech Excellence Program, where cloud computing services were taught as it relates to Big Data. The capstone project involved deploying an automated data pipeline using AWS CDK (Typescript flavor): loaded CSV & JSON data files to S3 Bucket (data lake), crawled the raw data using Glue Crawler, created tables and views in Athena for final visualization on QuickSight
  - **Skills**: *AWS-CDK, S3, Athena, QuickSight, Glue, Spark SQL, Lambda, Typescript*

- **Bloomberg L.P.** — Manhattan, NY
  Digital Transformation/Implementations & Support Associate — *Sep 2021 – Mar 2022*
  - **Collaborating with Cross-functional Teams**: Defined solution requirements to improve both client-facing and internal platforms by involving User Experience Engineering teams and other stakeholders
  - **User Research**: Conducted user research and generated analytical insights that were incorporated into a client-facing Q & A platform reducing the number of tickets generated
  - **Technical Support**: Provided technical guidance and solutions to more than 35 clients on how to generate new SSL certificates for continuous authentication to access Bloomberg financial data through API
  - **Skills**: *Python, Splunk, Jupyter Notebook, JIRA, Microsoft Office, Bloomberg Terminal*

- **Merck & Co. Inc.** — West Point, PA
  Data Engineering Intern — *Jun 2021 – Aug 2021*
  - **Automation**: Implemented a **scalable delta lake on top of the S3** data lake to make data available for consumption by downstream stakeholders using specified design patterns on Databricks
  - **Data Skills**: Developed an **end-to-end batch OLAP (online analytical processing) data pipeline** using data engineering best practices on Databricks with ELN (Electronic Laboratory Notebooks) data from AWS S3 bucket
  - **Technical Content Creation**: Documented and shared the data engineering process through **recorded video and written tutorials** in the Data Science Community of Practice forum for other teams to replicate and improve upon their pipelines
  - **Skills**: *Databricks, Spark, SQL, Data, Delta Lakes, AWS(S3), Confluence*

- **U.S. Army (Medical Research Institute of Infectious Diseases)** — Frederick, MD
  Genomics Research Supervisor & Soldier — *Jul 2014 – Sep 2021*
  - **Leadership**: Supervised and directed the daily workflow of 15 military and civilian personnel resulting inconsistent, greater than 98%, project completion, and quarterly department awards
  - **Organizational Skills**: Led and evaluated quality control within the laboratory leading to 2 successful biennial inspections by The Joint Commission and Center for American Pathologists
  - **Proactive**: Developed an improved genomics data processing pipeline to automate data transfer to the server using a Python script, increasing on-time delivery of data by over 90%
  - **Skills**: *Linux, Python*

EDUCATION

- **University of Pennsylvania** — Philadelphia, PA
  Master's Degree in Computer and Information Technology (part-time) — *2024 (Expected)*

- **Southern New Hampshire University** — Manchester, NH
  Bachelor of Science in Health Sciences — *Jun 2018 – May 2019*

- **George Washington University** — Washington, DC
  Associates' of Science in Health Sciences — *Feb 2008 – Nov 2011*