

Pharoah Evelyn

pharoahevelyn@gmail.com | (347) 439-7335 | [LinkedIn](#) | [GitHub](#) | [Blog](#)

I'm a Data Engineer with experience in Python, SQL, and the AWS cloud, possessing two active certifications, with expertise in machine learning, deep learning, and NLP - all of which can be found on my GitHub profile.

TECHNICAL SKILLS

Data Engineering: Apache Airflow, Apache Kafka, Data Pipelines, Data Warehousing, DBT, ETL, ELT • Amazon Web Services • Git •
Data Analysis: Apache Spark, NumPy, Pandas, Python, SQL • Data Visualization: Amazon QuickSight, Power BI, Tableau •
Machine Learning: Keras, Scikit-Learn, Statsmodels, TensorFlow • Natural Language Processing: NLTK • Linux •
Scripting: Ansible, BASH, Python • Web Scraping • Virtualization: AWS Management Console, VMware, VCenter, ESXI, Virtual Box

TECHNICAL PROJECTS

NYC Zillow Listings with Apache Airflow and Data Visualization - [Github](#)

Curation of a data pipeline that pulls data from a Zillow API into AWS using Apache Airflow with data visualization via Amazon QuickSight

- Collected Zillow listings in NYC from a Zillow API via Rapid API.
- Employed the Zillow API within Airflow to save listings within an S3 Bucket and transform data with Lambda functions.
- Visualize transformed data with Amazon QuickSight to view the best listings around the city based on geospatial data.

Data Engineering with Apache Airflow, Snowflake, and DBT with Cosmos - [Github](#) | [Demo](#)

Curation of an ETL pipeline using Apache Airflow, DBT, and Snowflake to manage collected data

- Employ Docker to create an airflow environment with Astronomer and its Cosmos package.
- Build a scalable pipeline using DBT within Airflow, centralizing data pipeline management and eliminating bash scripts.
- Snowflake warehouse was employed to receive the data from Airflow for analysis.

Serverless Spotify ETL Metadata Management - [Github](#)

Development of an ETL pipeline using AWS services to obtain the artist, album, and song information from my "Discover Weekly" playlist

- Collected JSON data through the Spotify API and ingested it into AWS with a Lambda function scheduled to run every week.
- Raw data is ingested into a raw zone, transformed into three separate tables, and then cataloged with AWS Glue.
- Cataloged data can be queried from AWS Athena, Redshift, or third-party apps like Big Query and Snowflake.

Real-Time Streaming Application with Apache Kafka - [Github](#)

Designing a real-time data processing engine to stream data for post-collection analysis

- Employed a managed Apache Kafka cluster to ingest real-time stock market data.
- Developed a Python notebook to auto-generate data every second to simulate real-time data ingestion.
- Transferred real-time data into storage to be queried within Athena in real-time for further analysis.

NYC Taxi Data Analysis with Apache Spark and Data Visualization - [Github](#)

Orchestrating a data pipeline with Apache Spark to transform NYC taxi data and perform data visualization via Amazon QuickSight

- Collected over 2,000,000 rows of parquet data to analyze taxi trips within NYC with incorporated location coordinates.
- Utilized Apache Spark to transform data into fact and dimension tables, preparing for data analysis with coordinates.
- Created a dashboard with Amazon QuickSight to discover insights based on new analytical data from SQL joins.

CERTIFICATIONS

[AWS Certified Data Analytics - Specialty](#)

10/2023

Amazon Web Services: Expertise in using AWS data lakes and analytics services to gain insights from data.

[AWS Certified Solutions Architect - Associate](#)

07/2023

Amazon Web Services: Ability to design cost-effective and performance-optimized solutions.

[AWS Certified Cloud Practitioner](#)

11/2019

Amazon Web Services: Validated foundational, high-level understanding of AWS Cloud, services, and terminology.

EDUCATION

Flatiron School, New York, NY

07/2021 – 07/2022

Immersive 600+ hour Data Science program

Full Sail University, Winter Park, FL

07/2017 – 05/2019

Bachelor of Science, Cloud Technologies

School of Visual Arts, New York, NY

09/2013 – 06/2017

Bachelor of Fine Art, Cartooning

EMPLOYMENT HISTORY

Deloitte Consulting LLP

Lake Mary, FL/Remote

Solution Analyst

07/2019 - Present

- Infrastructure Engineer for applications, hardware, virtualization, data, and websites used in government projects in the States of Georgia and Pennsylvania and the Exelon Corporation, using Linux OS and VMware ESXi via VCenter.
 - Oversaw and monitored the maintenance and troubleshooting of each project's software, hardware, and virtualization.
 - Collaborated with cross-functional teams across three projects to solve technical issues, control user access, and set guidelines to ensure compliance with project requirements.