Pranav Modh

hello@modhpranav.com | (857) 385 9511 | United States | /in/modhpranav | github.com/modhpranav | modhpranav.com

M.S. Computer Science | University of Massachusetts Boston | Graduation: May 2024 | GPA: 3.8 | Open to Relocate Anywhere in USA

Skills & Abilities

- **Programming Languages:** Python, Java, R, JavaScript, Unix/Bash Scripting
- **Data Platforms:** Snowflake, Palantir Foundry, DBT, DataBricks, Airflow
- **Big Data Technologies:** Apache Spark, Kafka, HDFS, Big Query, Redshift
- **Visualization Tools:** D3.js, Vega-Lite, Google Looker, PowerBI, Matplotlib
- **Libraries:** Spacy, NumPy, Pandas, Dask, Pyspark, FuzzyWuzzy
- **Frameworks:** Flask, FastAPI, SpringBoot, Django
- **CI/CD Tools/Networks:** Docker, Nginx, CircleCI
- **Databases:** PostgreSQL, MongoDB, MySQL, Oracle
- **Cloud:** AWS (EC2, S3, SQS, Lambda, Sagemaker, EKS), Google Cloud (Cloud Functions, Pub/Sub, Vertex AI, Data Storage, Compute Engine)
- **Machine Learning Certification:** https://www.coursera.org/account/accomplishments/certificate/M2G78BN8P9NX

Professional Experience

DATABASE DEVELOPER, PROVIDRS CARE, WICHITA KS **PRESENT**

- Developed a scheduled data pipeline to ingest data from **Excel, Microsoft Access, and Flat Files** into **AWS S3** for global storage.
- Improved data processing efficiency and reliability by implementing **Airflow DAGs on EC2** for data quality checks and ingestion into **Redshift**.
- Designed analytical schemas using **DBT** and integrated them with **Power BI**, enhancing reporting capabilities and **data-driven decision-making for stakeholders**.

DATA ENGINEER INTERN, MILLIPORESIGMA (MERCK KGAA), BOSTON MA **05/2023 – 12/2023**

- Led the **optimization of data workflows** in the semiconductor industry using **Palantir Foundry, PySpark, and TypeScript**, increasing data pipeline **efficiency by 45%** through advanced analytics and **AWS integration (EC2, S3, Lambda, SQS)**.
- Engineered data solutions with Redshift and DBT, **reducing processing time by 50%** and cloud resource **costs by 30%** through streamlined data transformation and governance.
- Developed a real-time data visualization dashboard using **Workshop Module and Google Looker**, enhancing stakeholder reporting and decision-making efficiency, and reducing response **times to insights by 25%**.

SOFTWARE ENGINEER, SIMFORM SOLUTIONS, INDIA **12/2021 – 07/2022**

- Spearheaded the development of a **multi-source data aggregation pipeline** using AWS S3, **Apache Kafka, and Apache Airflow**, reducing data **inconsistency by 40%** and improving data availability for analytics.
- Implemented automated ETL processes with **Snowflake and Unix/Linux scripting** to enhance data processing speed and **accuracy by 60%**.
- Established **CI/CD pipelines with CircleCI and Docker**, achieving a **75% improvement in deployment reliability**, and integrated Amazon EKS for scalable microservice management.

ASSOCIATE SOFTWARE ENGINEER, TNTRA, INDIA **08/2019 – 12/2021**

- Developed and deployed a web service for dynamic data handling using **Python, FastAPI, and MongoDB**, incorporating Unix scripting for automation, improving **system efficiency by 50%** and **reducing operational costs by 20%**.
- Led the integration of DevOps practices with **Docker, EKS, and GitHub** Actions, streamlining workflows, **reducing deployment cycles by 45%**, and enhancing software delivery efficiency.
- Built advanced analytics capabilities with **Apache Spark** and real-time data visualization using **Matplotlib and Vega-Lite**, enhancing data insight generation and user experience.

Projects (More Projects At: modhpranav.com/projects)

HEALTHCARE FRAUD DETECTION, Technology Used: Databricks, Pyspark, PostgreSQL, Aws S3, Aws Lambda, Data Modeling

- Leveraged **Databricks and PySpark** to process and analyze **U.S. healthcare claims data** from **Amazon S3**, optimizing a **PostgreSQL database** for enhanced data quality through sophisticated cleansing and aggregation.
- Developed a custom algorithm in Databricks to optimize claims processing and enhance **fraud detection**, utilizing PySpark for scalable analytics and pattern recognition. Integrated **Databricks with Amazon S3** to manage processed claims and **trigger AWS Lambda** for real-time alerts to stakeholders, improving decision-making and operational efficiency in healthcare claim management.

SUPPLY CHAIN OPTIMIZER, Technology Used: Palantir Foundry (ELT Tool), Java, AWS S3, AWS Lambda, Data Governance

- Utilized **Foundry's data integration tools** to process **Amazon S3 data** into a **relational Ontologized** database, enhancing data quality through aggregation and cleansing. Developed custom optimization algorithm within Foundry to optimize supply chain routes, leveraging **its built-in components for location-based calculations** and mappings between facilities.
- Integrated Foundry with Amazon S3 and **AWS Lambda** to manage and send optimized routes, improving decision-making efficiency.