Hemanth Pothineni

Dallas, TX | hemanthpothineni@outlook.com | 919 533 8762

https://www.linkedin.com/in/hp09/ | https://github.com/ph0923

Professional Summary

I'm a Data Engineer who builds scalable ELT and data infrastructure on the cloud, specializing in the modern stack, which includes Machine Learning, Databricks, Snowflake, Airflow, and dbt. My core value is translating raw data into business-ready assets, which I've achieved by optimizing query performance by up to 40% and improving data delivery speed by 35%. On the side, I've also leveraged these engineering skills to build an end-to-end real-time ML pipeline using Apache Kafka (Redpanda) for credit fraud detection, demonstrating my ability to handle a high-volume data problem from ingestion to live visualization in Streamlit.

Skills

- **Programming & Scripting:** Python (Pandas, NumPy, Scikit-learn, ML/Anomaly Detection), SQL (Postgres, MySQL, Oracle), Shell Scripting, PyCharm, Visual Studio
- **Data Warehousing & BI:** Snowflake, dbt (Data Build Tool), Azure Data Lake, Power BI, Tableau, Streamlit
- **Cloud Platforms & DevOps:** Azure, AWS, Azure DevOps (Azure Repos, CI/CD), Git, GitHub, GitLab, S3, ECS
- **ETL & Orchestration:** Apache Airflow, Azure Data Factory (ADF), Databricks (Apache Spark), SSIS, AWS Glue, Apache Kafka, Talend, Informatica
- **Databases & Tools:** PostgreSQL, Microsoft SQL Server, MySQL, Oracle
- **Operating Systems:** Windows, Linux, IOS
- **Methodologies & Practices:** SDLC, Agile, Waterfall, Data Modeling (Dimensional), Data Wrangling, Data Governance, ETL methodology, DevOps (CI/CD)

Experience

CBRE Group   **Sep 2025 – Present**

Data Engineer   **Dallas, TX**

- Designed and maintained robust ELT pipelines using Apache Airflow for orchestration and dbt (data build tool) for automated in-warehouse transformation, improving data delivery speed into Snowflake by 35%.
- Developed optimized SnapLogic integration workflows to ingest cross-system data, implementing rigorous validation logic that reduced data load errors by 25% through automated error handling.
- Engineered complex SQL queries and Python scripts to automate data cleaning, aggregation, and transformation, boosting query performance by 40% and eliminating manual intervention in recurring jobs.
- Architected dimensional data models in Snowflake and developed interactive Power BI dashboards, enabling 5+ business units to track real-time KPIs for data-driven decision-making.
- Managed CI/CD pipelines and version control using Azure DevOps and Git, streamlining the deployment of data solutions and reducing release cycle times by 30%.

Metamor Systems   **Apr 2024 – Aug 2025**

Data Engineer   **Indianapolis, IN**

- Migrated core ingestion processes from previous batch scripts to a hybrid Airflow/Azure Data Factory architecture, successfully addressing data latency issues and improving data availability speed by 15%.
- Utilized Databricks (Apache Spark) to handle the backlog of 3TB+ of historical case data, developing reliable processing jobs for quality monitoring and detailed agent error trend identification.
- Developed scikit-learn and Python preprocessing pipelines to structure eligibility data, automating the error classification process and reducing reliance on manual quality review by 40%.
- Automated manual reporting workflows by extracting processed data from the Azure Data Lake, streamlining delivery to OECOSL stakeholders and saving the team over 10 hours of manual preparation weekly.
- Tuned critical PostgreSQL queries and indexing strategies that underpinned caseworker queue management, resulting in a 45% increase in core application query response time.
- Implemented Python-based data anomaly checks within Airflow DAGs to proactively monitor ETL health, decreasing critical data-related pipeline incidents by 20%.

Maternal 911 Education Systems   **Aug 2023 – Dec 2023**
Business Intelligence Engineer   **Ithaca, MI**

- Engineered scalable analytics pipelines on Azure, utilizing Python for ETL processing and performing PostgreSQL database optimization to model complex patient care workflows.
- Delivered 25+ automated reporting solutions using advanced SQL Stored Procedures and Python scripts, providing critical visibility into healthcare KPIs and operational metrics.
- Developed interactive Power BI dashboards for clinical operations, ensuring data accuracy by implementing upstream Python validation scripts while optimizing data models for faster load times.
- Achieved a 25% increase in process accuracy and regulatory compliance by deploying SQL-based logic and Python automation to replace manual data tracking.
- Managed analytics request workflows via JIRA, executing ad-hoc SQL analysis to resolve tickets and maintaining clear technical communication with clinical and non-technical stakeholders.

Projects

Credit Fraud Detection by using ML

- Designed and implemented an end-to-end machine learning pipeline to secure financial transactions using unsupervised learning techniques.
- Architected a streaming data pipeline using Apache Kafka (Redpanda) to ingest and process high-volume transaction data with low latency.
- Reduced false positives by 20% through iterative tuning of the Isolation Forest algorithm, improving the reliability of the alert system.
- Developed a dynamic Streamlit dashboard refreshing every 5 seconds, visualizing live anomaly scores and transaction distributions for immediate monitoring.

Education

Master of Science in Operations Management and Information Systems   **Dec 2023**
Northern Illinois University   DeKalb, IL

Certifications

- Certified Power BI Data Analytics Associate-Microsoft
- Azure Administrator Associate-Microsoft
- Certified Tableau Professional