(561)-359-4473
ramthalla045@gmail.com

Ram Thalla

Python Developer - AI/ML

LinkedIn:RamThalla

Professional Summary:

- A **senior Python developer** and **data engineer** with more than **9 years** of expertise building scalable **data pipelines**, cloud-based data solutions, and **AI/ML** integrations across healthcare, energy, and enterprise domains.
- Demonstrated ownership of the full lifecycle for **AI-driven** clinical decision support modules, from conception to production deployment.
- **Engineered features** from complex clinical datasets to build predictive classification models for conditions such as diabetes and heart disease. Automated the training and deployment of these models using **Python scripts** to establish scalable and repeatable **ML workflows**.
- Validated model performance using key metrics including **precision, recall**, and **ROC-AUC**. Integrated model predictions into production **APIs** to support real-time, data-driven decision making for healthcare providers.
- Conducted **A/B testing** to empirically evaluate the impact of ML outputs on clinical decisions and developed custom logging tools to monitor model performance and **data drift** in production environments.
- Architected and implemented modern data solutions using **Microsoft Azure PaaS** services to facilitate real-time **data visualization** and reporting.
- Led the successful migration of legacy data systems to **Azure Synapse Analytics** and **Azure Data Lake Storage** , achieving a **25 percentage reduction in infrastructure** costs while improving scalability.
- Managed large-scale data ingestion from servers into **HDFS,** followed by bulk loading into **HBase** for scalable storage and retrieval. Administered and performance-tuned **Spark Databricks** clusters to ensure optimal resource utilization.
- Automated the deployment and scaling of cloud data services using **Azure Resource Manager** templates.
- Built and maintained robust, scalable **ETL pipelines** using **Python** and **Apache Airflow** to process high volumes of customer and transactional data from diverse sources.
- Automated manual **data ingestion workflows**, resulting in reduced processing times, fewer errors, and enhanced reliability.
- Identified and resolved **performance bottlenecks** within data pipelines to significantly improve **data throughput** and system reliability.
- Utilized **SQL** and **Tableau** to develop analytics for a marketing campaign, tracking **KPIs** and contributing to a **15 percent increase in client retention.**
- Authored and optimized complex **SQL queries** for dashboards and ad hoc reporting, achieving performance **improvements of up to 40 percent**.
- Designed and assisted in creating a centralized **data mart** using a **star schema** to unify disparate customer touchpoints into a single source of truth for analytics.
- Engineered and deployed **RESTful APIs** with **Python, Flask,** and **SQLAlchemy** to serve real-time clinical data and ML model outputs to healthcare dashboards.
- Managed the end-to-end build and release processes for multiple production modules within a **CI/CD** framework using **Visual Studio Team Services (VSTS).**
- Successfully l**ed cross-functional teams** and assumed full **ownership of project delivery** to ensure timely and high-quality outcomes.
- Actively **mentored junior engineers** in **Python development**, **data engineering** best practices, and **ML deployment** strategies, fostering team growth and technical excellence.
- **Collaborated** effectively with **data scientists, clinical analysts, and business stakeholders** to align technical solutions with strategic objectives and drive innovation.

Skills

Programming and APIs: Python, Flask, SQLAlchemy

Databases: PostgreSQL, PL/SQL

ETL and Data Engineering: Apache Airflow, Custom ETL scripts, Data cleaning, Transformation

AI/ML: Feature engineering, Classification models, Model evaluation, ML integration, A/B testing

Cloud Platforms: Microsoft Azure (PaaS), Azure Synapse Analytics, Azure Data Lake Storage

Big Data: HDFS, HBase, Spark, Databricks

DevOps and CI/CD: Visual Studio Team Services (VSTS), Azure Resource Manager

BI and Reporting: SQL, Tableau

Collaboration: Cross-functional teams, Technical documentation, Data profiling

Professional Experience

Senior Python Developer with AI/ML **June 2022 —PRESENT**

JPMorgan Chase *USA*

- Developed robust ETL pipelines to extract, transform, and load healthcare data from diverse sources into structured formats for analytics and machine learning workflows.
- Designed and maintained relational databases using PostgreSQL and PL/SQL, ensuring optimized performance and scalability for AI-driven analytics.
- Built and deployed RESTful APIs using Python, Flask, and SQLAlchemy to serve real-time clinical data and ML model outputs to healthcare dashboards.
- Engineered features from clinical datasets to support ML models for disease prediction, including diabetes and heart disease risk scoring.
- Collaborated with data scientists to train and validate classification models using structured patient data, contributing to improved risk stratification.
- Assisted in model evaluation by preparing validation datasets and analyzing metrics such as precision, recall, and ROC-AUC.
- Integrated ML model predictions into production APIs to support real-time decision-making for healthcare providers.
- Identified performance bottlenecks in data pipelines and implemented optimizations to improve throughput and reliability.
- Managed build and release processes for multiple modules in production using Visual Studio Team Services (VSTS).
- Designed and implemented data validation and preprocessing routines to ensure high-quality inputs for machine learning models used in clinical risk prediction.
- Automated the training and deployment of classification models using Python scripts, enabling scalable and repeatable ML workflows.
- Developed custom logging and monitoring tools for ML pipelines to track model performance and data drift in production environments.
- Led cross-functional collaboration efforts, mentoring junior developers and guiding data scientists on ML integration best practices to enhance hospital decision support systems.
- Conducted A/B testing on ML model outputs to evaluate impact on clinical decision making and refine model parameters based on feedback.
- Drove the business validation of AI solutions by integrating model predictions into clinical workflows and conducting rigorous A/B testing to measure their direct impact on provider decision-making.
- Mentored junior engineers on Python development and ML deployment strategies, fostering team growth and technical excellence.
- Took ownership of end-to-end delivery for AI-driven clinical decision support modules, ensuring timely and high-quality implementation.

Technologies Used:

- **Programming and APIs:** Python, Flask, SQLAlchemy.
- **Databases:** PostgreSQL, PL/SQL.
- **ETL and Data Engineering:** Custom ETL scripts, data cleaning, transformation.
- **AI/ML:** Feature engineering, classification models, model evaluation (precision, recall, ROC-AUC), ML integration, A/B testing.
- **DevOps and CI/CD:** Visual Studio Team Services (VSTS).
- **Collaboration:** Worked closely with data scientists, clinical analysts, and cross-functional teams.

Cloud Data Engineer **March 2019 — May 2022**
Apple *USA*

- Analyzed, designed, and built modern data solutions using Azure PaaS services to support real-time data visualization and reporting.
- Extracted large volumes of structured and unstructured data from servers into HDFS, followed by bulk loading into HBase for scalable storage and retrieval.
- Estimated cluster sizing and managed Spark Databricks clusters, including performance monitoring and troubleshooting to ensure optimal resource utilization.
- Developed and maintained CI/CD pipelines for multiple production modules using Visual Studio Team Services (VSTS).
- Implemented data ingestion workflows and transformation logic to support downstream analytics and business intelligence tools.
- Collaborated with cross-functional teams to align data architecture with business requirements and operational goals.
- Optimized data flow and storage strategies to reduce latency and improve throughput across distributed systems.
- Ensured data integrity and security across cloud environments by applying best practices in access control and encryption.
- Automated deployment and scaling of data services using Azure Resource Manager templates and scripting.
- Provided technical documentation and knowledge transfer sessions to support ongoing maintenance and onboarding.

Technologies Used:

- **Cloud Platform:** Microsoft Azure (PaaS).
- **Big Data:** HDFS, HBase, Spark, Databricks.
- **DevOps and CI/CD:** Visual Studio Team Services (VSTS), Azure Resource Manager.
- **Data Engineering:** Data ingestion workflows, transformation logic, performance optimization.
- **Security and Governance:** Access control, encryption.
- **Collaboration and Documentation:** Technical documentation, cross-functional team alignment

Data Engineer **March 2016 — Feb 2019**
DXC Technology *India*

- Built and maintained scalable ETL pipelines using Python and Apache Airflow to process customer and transactional data from multiple sources.
- Migrated legacy data systems to Azure Synapse Analytics and Azure Data Lake Storage, improving scalability and reducing infrastructure costs by 25 percentage.
- Developed optimized SQL queries for dashboards and ad hoc reporting, improving query performance by up to 40 percentage.
- Automated manual data ingestion workflows, reducing processing time and errors while improving reliability.
- Collaborated with BI and product teams to understand data requirements and deliver clean, well-structured datasets.
- Wrote technical documentation for pipeline architecture, data lineage, and recovery procedures.
- Collaborated on a marketing analytics project to track campaign KPIs using SQL and Tableau.
- Assisted in the creation of a centralized data mart using star schema to unify multiple customer touchpoints.
- Performed data cleaning and transformation using Python (Pandas) on CSV and Excel data from sales teams.
- Generated insights for client presentations, contributing to a 15 percentage increase in client retention for Q1 2022.

Technologies Used:

- **Programming and Scripting:** Python (Pandas)
- **Data Storage and Processing:** CSV, Excel
- **Data Warehousing:** Azure Synapse Analytics, Azure Data Lake Storage
- **ETL and Workflow Automation:** Apache Airflow
- **Databases and Querying:** SQL
- **Visualization and Reporting:** Tableau
- **Data Modeling:** Star Schema
- **Documentation:** Technical documentation for pipeline architecture and data lineage.

Education

B.Tech in Computer Science, *JNTUH* **Aug 2012 - May 2016**