Aditya Bhat

New Brunswick, NJ | LinkedIn| Github |aditya.bhat@rutgers.edu| +1 (732) 763-7131

SUMMARY

Python and SQL developer with 4 years in ETL, data automation, testing and project delivery using PySpark, Azure, AWS, and Airflow.

SKILLS

Language: Python, SQL, C++, C#, Shell scripting
Data Orchestration: Azure Data Factory, Azure Databricks, Apache Airflow
Cloud: AWS S3, EC2, Amazon RDS, Azure blob storage, Azure Data Lake Storage
Libraries: PySpark, Flask, Numpy, Pandas, Matplotlib, scikit-learn, PIL, ARCore.
Testing: Unit testing, A/B testing, Hypothesis testing

Visualization tools: Power BI, ggplot2, seaborn
Database: Postgres SQL, MySQL, Oracle dB.
Tools: Git, Docker, Unity3D
Editors: Pycharm, Visual Studio Code.

EDUCATION

MSc In Computer Science| Rutgers University, New Brunswick, NJ | CGPA: 4.0 **Sep 2021 - May 2023**
BE In Information Science| BMS Institute of Technology Bangalore, India **Aug 2013 - Jun 2017**

WORK EXPERIENCE

Nokia Bell Labs **New Jersey, US**
Data Science Intern **Jun 2022 - Aug 2022**

- Developed a robust ETL pipeline to process data over 100GB using **Pyspark**, with extensive **error handling** and **logging** to guarantee data integrity and traceability.
- Leveraged **Unix scripts** to trigger the pipeline and applied various transformation and compression techniques (bzip2, snappy) to optimize storage of vast datasets.
- Conducted **Unit testing** on transformed data using unit test libraries to validate script accuracy and logic implementation.
- Used **PostgreSQL database** to load the generated reports ensuring flexible, scalable, and efficient storage, facilitating easy querying and report generation.
- Collaborated with analytical team to produce critical analytical reports, ranging from city-specific analytics to prescriber patterns, enabling the client to refine and target their marketing strategy for a new product launch.
- *Skills Attained: Pyspark, Unix, HDFS, ETL, Postgress SQL database.*

Thinking Stack **Bangalore, India**
Data Engineer **Oct 2020 - Jun 2021**

- Automated the real-time processing of extensive COVID-19 datasets by setting up event-triggered pipelines, transferring data seamlessly from **Azure Blob** to **Azure Data Lake Storage Gen2.**
- Mounted Azure Data Lake onto **Databricks**, executing complex transformations using PySpark to extract meaningful patterns and trends from the COVID-19 data.
- Migrated transformed data to **Azure SQL** and leveraged **PowerBI** to craft dashboards, enhancing the Health Department's COVID-19 strategy and resource allocation.
- Developed an **Apache Airflow DAG** for an end-to-end pipeline, extracting and loading new papers from the National Library of Medicine to **AWS RDS**, hosted on **EC2**.
- Utilized Python scripts to download the latest papers to an **S3 bucket** and transformed XML data into a clean dataframe, automating the extraction and transformation process.
- Connected to **AWS RDS MySQL database** for iterative data loading and fully automated the ETL process, with daily execution controlled through a cron job in Apache Airflow, ensuring timely and consistent data updates.
- *Skills Attained: Azure Data Factory, Data Lakes, DataBricks, ETL, Postgress SQL, Apache airflow, ETL, AWS S3, EC2, RDS,*

DHS Informatics **Bangalore, India**
Machine Learning Engineer **Mar 2019 - Jul 2020**

- Developed 5 custom **YOLOv5** models for detecting safety PPE kit, and trained with up to 4 augmentation techniques achieving an overall mAP(.50) value of 0.81.
- Built an image augmentation pipeline that enhanced model performance through data processing and increased lower-class image counts, resulting in a 10% accuracy boost.
- Developed a POC for on-screen logo detection and analysis, leveraging **YOLOv5, pandas, and matplotlib** to assess attributes like screen time, size, and frequency.
- Constructed an interactive **Flask** interface for streamlined data visualization in advertisement logo analytics.
- *Skills Attained: YOLOv5, Pandas, Matplotlib, Flask.*

Oracle Financial Service Software **Bangalore, India**
Technical Analyst **Jun 2017 - Mar 2019**

- Extracted and organized customer transaction data from Oracle Financial Crime and Compliance Management using complex SQL queries and joins.
- Cleaned data by handling missing or erroneous values with SQL functions such as COALESCE and NULLIF.
- Analyzed customer spending patterns and transaction frequencies using SQL functions (SUM, COUNT, AVG) for aggregated metrics, and employed SQL window functions (RANK, PARTITION BY) to identify top spenders and frequent transaction patterns.
- *Skills Attained: SQL Query Optimization, Data Cleansing, Data Analysis, Automation, Debugging, Client Coordination.*