Muhammad Abdullah Nasir

mnasir3@gsu.edu | (470) 301-1086 | linkedin.com/in/m-abdullah-nasir/ | github.com/daredevilx616 | daredevilx616.github.io

EDUCATION

Georgia State University — Atlanta, GA **December 2024, Expected May 2026**
Bachelor of Science in Computer Science, Master of Science in Computer Science *GPA: 3.85*

EXPERIENCE

The Coca-Cola Company **June 2024 – August 2024**
Data Engineering Intern *Atlanta, GA*

- Orchestrated media data ingestion from Dentsu Inc., leveraging Scala and Spark within a Medallion ETL architecture to efficiently transition data through raw/bronze, refined/silver, and certified/gold zones, improving data integrity and reducing processing time by 25%.
- Employed Azure Synapse to construct and automate data pipelines, ensuring seamless data ingestion and integration, reducing manual intervention by 30%.
- Conducted comprehensive semantic and structural checks on over 500 CSV files in the raw zone, converting them to delta file format in the refined zone, and applying necessary transformations in the certified zone, increasing data accuracy by 20%.
- Engineered robust data processing workflows using Spark SQL to convert and store transformed data into 10+ SQL tables, subsequently loading them into the data warehouse, optimizing data accessibility and utilization.
- Developed and executed unit tests and test cases for each zone, class, and method, ensuring robust data validation and quality control throughout the ETL process.

The Coca-Cola Company **May 2023 – May 2024**
Data Engineering Co-Op *Atlanta, GA*

- Streamlined data workflows through DevOps and Azure Data Factory, collaborating with Databricks to automate ETL (Extract, Transform, Load) processes, including data ingestion, cleansing, filtering, and calculations, achieving a 20% reduction in data processing time.
- Utilized Azure Data Factory to design and deploy data pipelines and ensuring seamless data flow, thereby enhancing operational efficiency, and reducing manual intervention.
- Integrated Workday SOAP API, efficiently parsing XML data using Python and Pyspark, optimizing data retrieval of company's over 30,000 employees and enhancing data management processes.
- Employed advanced Python and SQL to extract and transform big data from SAP tables/APIs, empowering data-driven decisions, and insights, contributing to a 25% increase in data accuracy.
- Executed complex MDX queries and managed data cubes proficiently, generating tailored, actionable reports for business users, debugging data issues, and reducing data-related errors by 30%.
- Pioneered Microsoft Azure Logic App, triggering prompt email notifications for subscribed users to resolve data table discrepancies, showcasing hands-on Azure technology expertise.
- Crafted dynamic Power BI visualizations and dashboards, effectively communicating complex insights to stakeholders, enhancing data comprehension, and aiding strategic decision-making.

PROJECTS

Social Media Analytics | *AWS S3, Glue, Redshift, Scala, Spark, Terraform, GitLab* **August 2024**

- Built a data pipeline on AWS for real-time sentiment analysis of social media data from multiple platforms.
- Developed Spark jobs in Scala for data transformation, reducing processing time by 35%.
- Automated CI/CD pipeline with GitLab, improving ETL deployment efficiency by 40%.

LeadLess – GT Hacklytics 23' | *Python, Matplotlib, Pandas, Jupyter Notebook, Folium, Geopy* **February 2023**

- Developed a website to raise awareness about the dangers of lead poisoning.
- Collected various datasets and performed data analysis and visualizations using Pandas, Mathplotlib and Folium libraries to back the argument that people must know about lead poisoning.
- Created line graphs, bar graphs, and choropleth maps to depict lead concentrations and emissions.

TECHNICAL SKILLS

Programming Skills: Java, Python, Scala, SQL, C++, HTML, CSS, JavaScript, React.Js, Dart
Developer Tools: Azure Data Factory, Azure DevOps, Azure Synapse, AWS Glue, AWS Redshift, S3, Data Lake/Warehouse, Apache Spark, Hadoop, Logic App, Flutter, Git/GitHub/GitLab, Rest/Soap APIs, MySQL, PostgreSQL
Libraries: PySpark, Selenium, Flask, Pandas, Tweepy, Matplotlib, BeautifulSoup, Folium, Numpy, Geopy, JavaFX