

BISWAJITH YETUKURI

Data Engineer

St. Charles, MO

(314)-467-8077 | biswajith.yetukuri.data@gmail.com | LinkedIn: [linkedin.com/in/biswajith-y/](https://www.linkedin.com/in/biswajith-y/)

PROFESSIONAL SUMMARY

Data Engineer with **4+ years of experience** building scalable data pipelines and cloud-based data solutions. Proficient in **Python, SQL, PySpark, and Databricks** for data ingestion, transformation, and processing. Experienced in developing reliable **ETL/ELT workflows** and maintaining high-quality, analytics-ready datasets. Strong background in delivering efficient data solutions for **business intelligence and analytics**.

TECHNICAL SKILLS

Programming:	Python, SQL, R, Java, JavaScript
Data Engineering:	ETL/ELT, Data Pipelines, Data Integration, Data Ingestion, Data Transformation, Data Modeling, Data Warehousing, Data Quality, Data Validation, Schema Validation
Data Processing:	Apache Spark, PySpark, Databricks, Delta Lake, Pandas, NumPy, dbt
Orchestration:	Apache Kafka, Apache Airflow, Spark Structured Streaming
Cloud:	Azure Data Factory, Azure Databricks, Synapse Analytics, ADLS Gen2, Azure SQL, Azure Functions, AWS S3
Data Warehouses:	Snowflake, Azure SQL, PostgreSQL, MySQL, MongoDB
DevOps:	Azure DevOps, CI/CD, Git, GitHub, Terraform, Docker, Kubernetes
Integration:	FastAPI, REST APIs, JSON, XML
Analytics:	Power BI, Alteryx, Tableau

PROFESSIONAL EXPERIENCE

Data Engineer PNC Financial

April 2025 – Present

- Designed **ADF and Synapse ETL/ELT pipelines**, integrating data from multiple source systems into ADLS Gen2 to support reliable downstream analytics and enterprise reporting.
 - Built and optimized **PySpark pipelines in Azure Databricks**, reducing pipeline execution time by **20%** through efficient transformation and partitioning improvements.
 - Developed **ADF Mapping Data Flows and REST API integrations** to cleanse, transform, standardize, and load curated datasets into Azure SQL Database for downstream analytics.
 - Implemented automated **data-quality and validation checks**, identifying inconsistent records while improving accuracy, consistency, and reliability across enterprise analytics datasets.
 - Automated **ADF and Databricks deployments using Azure DevOps CI/CD**, improving deployment consistency while supporting version control, monitoring, and reliable pipeline releases.
-

Data Analyst Intern Husmann Corporation

May 2024 – September 2024

- Developed **Python-based data processing workflows on AWS**, automating refrigeration dataset comparisons and reducing manual verification effort by approximately **20%**.
- Cleaned and transformed **refrigeration sensor and leak datasets**, preparing reliable, analytics-ready data for engineering research and downstream predictive modeling.
- Performed **exploratory data analysis on sensor datasets** to identify patterns, anomalies, and data inconsistencies supporting refrigeration-system analysis and engineering decisions.
- Implemented automated **data-quality checks** for missing values, duplicates, inconsistencies, and anomalous sensor readings before downstream analytics and reporting.
- Integrated processed datasets into **automated reporting workflows**, improving accessibility of refrigeration insights and enabling more efficient analysis across cross-functional engineering teams.

- Developed scalable **batch and streaming pipelines using Apache Spark, PySpark, Python, and SQL**, supporting high-volume operational data processing for enterprise analytics and business reporting.
 - Built reusable **data transformation and cleansing workflows**, reducing data-preparation effort by **20%** while improving consistency, accuracy, and reliability of downstream analytical datasets.
 - Designed **REST API-based ingestion pipelines**, integrating data from **5+ third-party sources** into centralized SQL databases to support downstream analytics and reporting workflows.
 - Optimized **Spark and SQL data-processing workflows**, reducing recurring pipeline execution time by **15%** while improving performance and scalability across operational data-processing workloads.
 - Implemented automated **data-quality checks, role-based access controls, and data masking**, improving dataset reliability while supporting enterprise data-security, governance, and compliance requirements.
-

PROJECTS

Real-Time Data Processing Lakehouse | [GitHub](#)

- Built an end-to-end **real-time data lakehouse** using Python, Kafka, PySpark, Spark Structured Streaming, Delta Lake, and Airflow, implementing **Bronze–Silver–Gold architecture** for scalable batch and streaming data processing.
- Implemented automated **data quality, schema validation, orchestration, and monitoring**, with Dockerized pipelines supporting both generated and user-provided datasets.

Cloud Data Warehouse & ELT Platform

- Built an end-to-end **ELT platform using Python, Airflow, AWS S3, Snowflake, and dbt**, automating multi-source ingestion and transforming raw data into analytics-ready models.
 - Implemented **dbt testing, incremental models, data-quality validation, and Airflow orchestration**, with Docker and CI/CD supporting reliable and reproducible pipeline deployments.
-

EDUCATION

Master's Degree (MS) - Data Analytics and Engineering
George Mason University

Aug 2023 – May 2025