HARISHANKAR GIRI
Data Engineer

📞 +91-8260991535 ✉ hsgiri.data@gmail.com in linkedin.com/in/hsgiri ⌨ github.com/HSG15

PROFILE

Data Engineering enthusiast with 1 year of startup experience and hands-on work in **SQL**, **Python**, **PySpark**, **Azure Data Factory**, **ADLS Gen 2** and **Databricks**. Skilled in building ETL pipelines, performing data cleaning and transformations, and working with cloud-based storage and processing tools.

TECHNICAL SKILLS

Programming & Querying: C++, Python, PostgreSQL
Big Data Technologies: Spark, PySpark, Spark SQL, Hadoop, HDFS
Cloud Computing: Azure Data Factory, Databricks, ADLS Gen2
Version Control & Tools: Git, GitHub, Postman, Docker

EXPERIENCE

Associate Technical Consultant   **Aug 2024 – Present**
Blue Flame Labs Pvt Ltd   *Pune, India*

- Designed and automated client data pipelines in **Azure Data Factory**, ingesting data via HTTP requests and storing it in NetSuite's File Cabinet, reducing manual uploads by 80%.
- Built scalable PySpark jobs in **Azure Databricks** to clean and transform raw data by handling null values and standardizing schemas, improving data quality by 35%.
- Converted refined CSV datasets into JSON objects and integrated them with NetSuite using the **SuiteTalk REST API**, reducing import errors by 40% and accelerating master record creation.
- Collaborated with stakeholders on schema validation and NetSuite integration, improving onboarding turnaround time by 20%.

PROJECTS

Azure E-commerce Data Pipeline | *Data Factory, ADLS Gen2, Databricks (PySpark), Synapse*   **July 2025**

- Worked with the Brazilian e-commerce dataset **Olist** (Kaggle), consisting of 9 interrelated tables with over **100k+ records**, simulating real-world data engineering workloads.
- Implemented a cloud-based **Medallion Architecture** (Bronze → Silver → Gold) using Azure services to process large-scale e-commerce data efficiently.
- Built dynamic ingestion workflows in **Azure Data Factory** to extract data from GitHub, MySQL, and MongoDB and load it into the Bronze layer of **ADLS Gen2**.
- Applied PySpark transformations in **Azure Databricks** and used **Azure Synapse CETAS** to deliver analytics-ready datasets to the Gold layer, improving query performance by 25%.

NASA APOD ETL Pipeline | *Apache Airflow, Docker, PostgreSQL, AWS RDS, Astronomer, Python*   **Aug 2025**

- Built an automated ETL pipeline in **Apache Airflow** to fetch daily images and metadata from **NASA's APOD API** and load them into **PostgreSQL**.
- Developed Python-based Airflow DAGs with task dependencies for API extraction, table creation, and data ingestion.
- Containerized the pipeline using **Docker Compose** to run Airflow and PostgreSQL locally, configuring HTTP and Postgres connections via the Airflow UI.
- Deployed the pipeline to **Astronomer Cloud** and connected it to **AWS RDS PostgreSQL**, enabling persistent storage and historical data retrieval.

CERTIFICATIONS

- **Databricks Fundamentals Accreditation**
- **Get Started with Databricks for Data Engineering**

EDUCATION

B.Tech. in Computer Science and Engineering   **2020 – 2024**
GIET University, Gunupur   *CGPA: 8.5 / 10*