

# BISWAJEET SAHU

## Data Engineer/Analyst

Phone: [+91-8789827182](tel:+91-8789827182), Email: [biswajeetsahu94@gmail.com](mailto:biswajeetsahu94@gmail.com), [LinkedIn Profile](#)

*Data Engineer with 2 years of experience in building scalable Data Pipelines, Datawarehouse solutions, ETL and ELT, Big Data, PowerBI dashboards and managing operations of multi-Terabyte systems for a world's leading air conditioning manufacturing client.*

---

## SKILLS

- **Languages:** Java, Python, SQL
- **AWS:** EC2, S3
- **Big Data & ETL:** HDFS, PySpark, Hive, YARN, SparkSQL, Airflow, CICD Pipelines, Snowflake, Snowpipe
- **Repository Management:** Azure Devops
- **Project Management:** JIRA
- **Orchestration:** CA Autosys
- **Analytics:** KNIME, PowerBI
- **Databases:** MSSQL, H2, MySQL, PostgreSQL, Graph DB like Oracle NoSQL DB
- **Visualization:** Excel, PowerBI
- **Deployment:** Docker, Kubernetes
- **Math:** Statistics, Calculus, Probability, PNC, Matlab, Linear Algebra

## EXPERIENCE

Jan 2022- July 2023

### **DATA ENGINEER** at **HCL TECHNOLOGIES**

- **Data Ingestion and Analysis**
  - Data Ingestion pipeline to ingest flat file into **Data Lake**
  - Automated flat file loaded into **AWS S3** to **Hive** table with the help of Orchestration tool like **CA Automation**
  - Designed **PowerBI dashboards** based on Data Analysis results.
- **ELT and Analysis**
  - Loading of metadata file from **AWS S3** bucket to **Snowflake** stage automatically using **Snowpipe**
  - Transformation of loaded data and analyze it using **SQL** queries
  - Visualization of the analyzed data using **PowerBI**
- **Data Migration**
  - Designed and built a data Transformation process using **KNIME, H2**
  - **Automated** workflows written in **KNIME** using batch file
  - Lead and mentored a member for bug fixing and designing new workflows as required

**Oct 2020-June 2021**

**Data Analyst at FSNL**

- Design, build, and maintain efficient **data pipelines** for ingesting, processing, and transforming data from various sources.
- Utilize Python to create custom **scripts** and tools for data processing and manipulation.
- Manage and optimize data in **Google BigQuery** to ensure accuracy, performance, and security.
- Develop strategies for **ingesting data** from files and other structured or unstructured sources.
- Work with **Cube Flow** to create data cubes and enhance data querying performance.
- Design and schedule data workflows using **Apache Airflow** for automation and orchestration.
- Handle secure data transfers via **SFTP** protocols
- Implement **data quality checks** and validation processes to maintain data accuracy and consistency.
- Create and maintain technical **documentation** for data pipelines, schemas, and processes.
- **Collaborate** with data scientists, analysts, and stakeholders to understand data requirements and deliver solutions.
- Identify and resolve performance bottlenecks to **optimize** data processing.

**EDUCATIONS:**

- **2014 – 2018**  
**BTECH , Mizoram University, Electrical Engineering - 8.73 GPA**
- **2011-2013**  
**High School from Kendriya Vidyalaya Surda - 87.80%**
- **2010-2011**  
**Intermediate from St. Joseph Convent High School - 87.14%**

**CERTIFICATIONS:**

- Spark Beyond basics(Udemy)
- MasterSQL for Data Science(Udemy)
- Snowflake (Udemy)