ASPARSH RAJ

Jamui, Bihar

📞 +91-7667406415 ✉ rajasparsh19@gmail.com 🌐 Portfolio in LinkedIn ⦿ GitHub Ⓗ Hackerrank 📊 LeetCode

Education

Sarala Birla University   **Sept, 2022 – Aug, 2025**
*BCA - **CGPA - 8.7***   *Ranchi, India*

Experience

Claw Legaltech   **Dec 2024 – Mar 2025**
Data Engineer Intern   *Ahmedabad, India*

- Developed and optimized **DBT data pipelines**, improving **ETL efficiency by 40%** and ensuring **scalability for 1 million+ data records** in the data warehouse.
- Implemented Git best practices (branching, merging, commit history tracking), reducing deployment errors by 30% and improving team collaboration.
- Built **DBT models** across staging, intermediate, and mart layers, contributing to the **end-to-end data pipeline development**.
- Optimized production DBT models by identifying redundant code and implementing **DBT macros**, **reducing development time by 78%** and **lines of code by 95%**, leading to a **2x improvement in team efficiency**.
- Implemented **DBT tests** on both **source and mart models** to **improve data quality and reliability**, minimizing the risk of data failures for downstream users.
- Automated the entire **data pipeline** to run in a **production environment on Databricks Cloud**, orchestrated using **Apache Airflow**.

BharatTech   **Apr 2024 – Sept 2024**
Backend/Data Engineer Intern   *Lucknow, India*

- Worked as both a **Data Engineer** and **Backend Engineer** intern, contributing to critical data migration projects and backend development, helping the organization scale efficiently.
- Led the migration from **NoSQL (Firebase Firestore)** to **SQL**, reducing query complexity and backend overhead by **40%**. This transition also reduced cloud costs by **70%** by minimizing reliance on Firebase.
- Utilized **PySpark** to clean, transform, and convert **semi-structured** data into **SQL**, improving data accessibility and analysis efficiency.
- Built and optimized **RESTful APIs** using **Django Rest Framework**, processing more than 10,000 API calls per hour, **improving data retrieval speeds by 50%**

Projects

dbt-Airflow Data Pipeline ⧉ | Tech Stack: **DBT, DLT, SQL, Airflow, Python, Git**   **Mar, 2025**

- Developed an **end-to-end data pipeline** to automate weather data ingestion and reporting, demonstrating proficiency in **DBT, DLT, and Apache Airflow** by successfully processing **hourly weather data** into structured **weekly reports**.
- Ingested **real-time weather data** from an external API and stored it in **MySQL using DLT**, ensuring **efficient and scalable** data extraction and loading.
- Transformed **raw weather data** into actionable **weekly insights** by building **DBT models**, improving **data accessibility and analysis**.
- Orchestrated the **entire workflow** using **Apache Airflow**, enabling **seamless scheduling, execution, and monitoring** of weekly data processing tasks, ensuring **automation and reliability**.
- Implemented **version control** in both the **DBT model** and **DLT ingestion pipeline** using **Git**, allowing for **efficient tracking of system changes and updates**.

Skills

Programming Languages: Python, SQL, Bash
Libraries/Framework: DBT, DLT, Apache Airflow, PySpark, Django
Cloud Platforms: Databricks
Databases & Tools : MySQL, Firebase, Git, Linux

Certifications

- SQL Advanced Certificate - Hackerrank, 2025
- PySpark - Apache Spark Programming in Python for Beginners - Udemy. 2024
- Google Cloud Data Engineer Professional Certificate - Coursera, 2023