

ASPARSH RAJ

Jamui, Bihar

☎ +91-7667406415

✉ rajasparsh19@gmail.com

📁 [Portfolio](#)

🌐 [LinkedIn](#)

🐙 [GitHub](#)

🏠 [Hackerrank](#)

📄 [LeetCode](#)

EDUCATION

Sarala Birla University

BCA - CGPA - 8.7

Sept, 2022 – Aug, 2025

Ranchi, India

EXPERIENCE

Neos Group

September, 2024 – Present

Web Scraping Data Engineer

Hyderabad, India

- Spearheaded **data scraping** project spanning **120+ countries** and **200 cities**, extracting data from Google Maps, resulting in **45% reduction in data gathering time**.
- Developed **Python scripts** that automated the extraction and processing of data from multiple online sources, using tools like **Playwright** and **BeautifulSoup**.
- Implemented **parallel processing** techniques to handle high volumes of data and ensure timely completion of data scraping tasks, which boosted performance by **30%**.
- Maintained a **clean and organized database** of all gathered data, ensuring seamless data retrieval and integration for downstream analysis and reporting.
- Implemented a robust **logging** mechanism of the entire **ETL** process to maintain the trackability of the framework

BharatTech

April, 2024 – September, 2024

Backend/Data Engineer Intern

Lucknow, India

- Played as both **Data Engineer** and **Backend Engineer** intern, contributing in critical data migration projects and backend development, helping the organization scale efficiently in a fast-paced startup environment.
- Led the migration from **NoSQL (Firebase Firestore)** to **SQL**, creating a complete **ETL pipeline**, reducing query complexity and backend overhead by **40%**. This change cut cloud costs by **70%** due to reduced reliance on Firebase.
- Collected and processed over **25 GB** of **semi-structured** data from disparate sources using **web scraping** tools such as **requests**, **BeautifulSoup**, and **Scrapy**, ensuring data consistency and integrity before storing it in on-premise filesystems.
- Used **PySpark** to clean, transform, and convert **semi-structured** data to **SQL**, enhancing data accessibility and analysis efficiency.
- Developed **RESTful APIs** for the transformed data using **Django Rest Framework**, now managing thousands of **API calls per hour** and providing valuable data-driven insights for platform optimization.

PROJECTS

PySpark app for Batch Data Processing 📄 | Tech Stack: Python, PySpark, Databricks

Oct, 2023

- Designed and implemented a data pipeline for weather data using Python, PySpark, and Databricks, achieving a **43% reduction** in manual data transformation time.
- Processed and analyzed over seven days of hourly weather data, visualizing insights with **Power BI dashboards** for improved decision-making.

SKILLS

Languages: Python, SQL

Developer Tools: VS Code, Git, GitHub

Technologies/Frameworks: PySpark, Django, Scrapy, PowerBI, MySQL, ETL, Pandas, BeautifulSoup, Requests, Web Scraping, Playwright

Coursework: Data Structures & Algorithms, Software Engineering, Database Management System (DBMS)

Other Skills: Problem Solving, Analytical Thinking, Written Communication, Verbal Communication

CERTIFICATIONS/COURSES

- PySpark - Apache Spark Programming in Python for beginners -Udemy
- Google Cloud Data Engineer Professional Certificate - Coursera
- IBM Data Engineering Professional Certificate - Coursera