

Rahul P

+91 8431053364 | rahuljobsearchhh@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | Bangalore, India

SUMMARY

Data Engineer with a strong foundation in Python, SQL, distributed data frameworks, and cloud architecture. Experienced in designing robust ETL/ELT pipelines, converting complex unstructured data into analytics-ready schemas, and optimizing system workloads. Adept at applying system design thinking and cost-optimization strategies to deliver scalable, reliable data infrastructure.

EXPERIENCE

Build-Ora Aug. 2025 – Present
Co-Founder Bangalore, India

- Led cross-functional technical teams to architect and execute major civic-tech events, including Hackademia 2025, establishing robust collaborative infrastructure and project management pipelines
- Spearheaded technical problem-solving initiatives, mentoring 50+ developers on backend engineering, data modeling, and system design thinking while translating complex architectural challenges into structured frameworks

EVAP Solars Oct. 2025 – Dec. 2025
Data Engineering Intern Bangalore, India

- Designed batch ETL pipelines and implemented validation logic using Python and SQL to migrate raw operational logs into optimized dimensional schemas (Star/Snowflake), increasing downstream data consistency by 25%
- Translated complex KPI metrics into high-performance analytical queries while enforcing automated pipeline testing to catch schema drift, ensuring data quality and significantly decreasing dashboard latency

PROJECTS

RANGO | *PySpark, Databricks, Apache Airflow, FastAPI, AWS S3* 2026

- Architected a scalable platform to benchmark Retrieval-Augmented Generation (RAG) datasets, building an ELT pipeline with Apache Airflow to safely ingest large volumes of unstructured inputs into AWS S3
- Engineered distributed data transformations using PySpark on Databricks via a Medallion Architecture, ensuring strict pipeline idempotency while optimizing parsing efficiency by 40% across parallel cluster nodes

Study Buddy | *Python, dbt, PostgreSQL, Docker, LangChain, Streamlit* 2026

- Developed a Docker-containerized orchestration pipeline that ingests complex unstructured PDFs to automatically parse, clean, and output structured study notes
- Integrated dbt to build structured relational tables within PostgreSQL, enforcing automated data quality assertions (null/uniqueness checks) to secure output integrity across testing and production environments

EDUCATION

The National College, Jayanagar Aug. 2023 – May 2026
Bachelor of Computer Applications – CGPA: 9.1 Bangalore, India

SKILLS

Languages & Core Frameworks : Python, SQL (CTEs, Window Functions), C, JavaScript

Data Engineering & Cloud : Apache Spark (PySpark), Databricks, Apache Airflow, dbt, AWS (S3, Redshift, Glue), ETL/ELT Pipelines

Databases & Storage : PostgreSQL, MySQL, Delta Lake, Star/Snowflake Schemas, Medallion Architecture

Tools & Practices : Git/GitHub, Docker, CI/CD Baselines, System Design, Cloud Cost Optimization, Agile/Scrum