Dhairya Pratap Singh Gaur

+91 7454948411 | dhairyagaur98@gmail.com
LinkedIn | **GitHub**

PROFESSIONAL SUMMARY

Data Engineer with 3+ years of experience building scalable **ETL** pipelines, big data processing, and cloud-based solutions. Proficient in **Python, SQL, PySpark, Databricks, and Snowflake**, with experience optimizing workflows and improving data quality. Passionate about delivering high-performance data solutions that drive business insights.

TECHNICAL SKILLS

Programming Languages:	Python, SQL
Big Data Technologies:	Spark, PySpark, Spark SQL
Cloud Platforms:	Databricks, Snowflake, Google Cloud Platform (GCP)
Data Engineering Concepts:	ETL, Data Warehouse, Data Lake, Data Lakehouse, Data Modelling
Visualization Tools:	Amazon QuickSight
Other Technologies:	pgAdmin, Git/Github, Airflow
Data Quality & Governance:	Data Validation, Data Governance

WORK EXPERIENCE

KPMG   November 2021 - Present
Data Engineer   *Bangalore, India*
- Optimized SQL scripts from Teradata for catalog tables to ensure efficient and effective migration.
- Built metadata scripts and constructed **ETL** data pipelines within the **Medallion Architecture** framework.
- Implemented data processing and transformation logic using **PySpark** and optimized queries with **Spark SQL**, reduced query runtime by **60%**.
- Deployed and managed scalable data solutions leveraging **Databricks** for collaborative development and performance tuning.

- Designed and implemented detailed reports and data analysis pipelines, increased data scalability by **30%**.
- Developed interactive, data-driven dashboards utilizing **Amazon QuickSight** for visualization.
- Managed database operations and optimized queries using **pgAdmin**.

SmartBridge   May 2020 - June 2020
Machine Learning Engineer (Intern)   *Hyderabad, India*
- Developed an automated AI system to predict individual life expectancy, achieved a predictive accuracy of **94%**.
- Incorporated various factors such as demographic, health, lifestyle, and socioeconomic data.
- Utilized IBM frameworks, including :
- **IBM Cloud** for infrastructure and scalability.
- **IBM Watson** for data processing and machine learning model development.
- **IBM Machine Learning Instance** for model training, deployment, and real-time predictions.

CERTIFICATIONS

- Databricks Lakehouse Fundamentals
- Databricks AWS Platform Architect
- Databricks Azure Platform Architect
- Databricks GCP Platform Architect
- Databricks Platform Administrator

EDUCATION

Bachelor of Technology (B.Tech) - Electronics and Communications   CGPA - 7.1
Guru Gobind Singh Indraprastha University   *August 2017 - July 2021*

Senior Secondary   12th - 77.4%
Kendriya Vidyalaya Sangathan   *April 2015 - March 2016*